Biosynthesis and Incorporation of Nonproteinogenic Amino Acids into Non-ribosomal Peptide Natural Products

Author: Paul Fredrick Widboom

Boston College

The Graduate School of Arts and Sciences

Department of Chemistry

BIOSYNTHESIS AND INCORPORATION OF NON-PROTEINOGENIC AMINO ACIDS INTO NON-RIBOSOMAL PEPTIDE NATURAL PRODUCTS

A dissertation

by

PAUL FREDRICK WIDBOOM

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

December 2008

Biosynthesis and Incorporation of Nonproteinogenic Amino Acids into Non-ribosomal Peptide Natural Products

Prof. Steven D. Bruner — Paul Fredrick Widboom

Research Advisor — December 8, 2008

Abstract

Complex and unique enzymology is often behind the biosynthesis of natural products. This thesis is focused on how non-proteinogenic amino acids are biosynthesized and then incorporated into natural products. Chapters two, three and four deal with a unique dioxygenase found in vancomycin biosynthesis. Chapter five elaborates on the biochemical characterization along with efforts toward structural characterization of a terminal non-ribosomal peptide synthetase module.

The vancomycin biosynthetic enzyme DpgC belongs to a small class of oxygenation enzymes that are not dependent on an accessory cofactor or metal ion. The detailed mechanism of cofactor-independent oxygenases has not been established. We have solved the first structure of an enzyme of this oxygenase class complexed with a bound substrate mimic. The use of a designed, synthetic substrate analog allows unique insights into the chemistry of oxygen activation. The structure confirms the absence of cofactors, and electron density consistent with molecular oxygen is present adjacent to the site of oxidation on the substrate. Molecular oxygen is bound in a small hydrophobic pocket and the substrate provides the reducing power to activate oxygen for downstream chemical steps. Our results resolve the unique and complex chemistry of DpgC, a key enzyme in the biosynthetic pathway of an important class of antibiotics. Mechanistic

parallels exist between DpgC and cofactor-dependent flavoenzymes, providing information regarding the general mechanism of enzymatic oxygen activation.

Acknowledgements

I would like to thank the members of my thesis committee, Mary Roberts, Marc Snapper and Jianmin Gao, for taking time to critically read this thesis, and to attend my defense. I also would like to thank my committee for a stimulating discussion during my private thesis defense. I also acknowledge those that read this thesis and gave constructive criticism: Heather Cooke, Elisha Fielding, Timothy Montavon, Eric Dimise Donna Payne and Janet Widboom.

I am indebted to my advisor, Steve Bruner, for all his patience and hard work in training me to become the scientist that I am. Steve expects nothing but the best from his graduate students and provides everything he can to make us successful. He has a passion for science and a curiosity for complex problems and questions in the field of biological chemistry. He has also had an open door for questions and assistance whenever I was in need.

All the members of the Bruner lab provided support to me during graduate school. Elisha Fielding worked side by side with me on the DpgC project. She established crystallization conditions for DpgC with Dpa-NH-CoA and worked tirelessly with me on biochemical assays. Ye Liu developed the synthesis of Dpa-NH-CoA, making it quick and easy to produce that invaluable small molecule. Eric Dimise's structural work on fuscachelin A made it possible to biochemically characterize FscI. Deborah Mitchell crystallized FscI and gave us a chance to solve the crystal structure. Heather Cooke, Timothy Montavon, Carl Christianson, Heather Condurso and Becca Goldstein all provided critical discussion covering the projects I worked on in graduate school.

The staff of the National Synchrotron Light Source at Brookhaven National Labs was instrumental in helping collect X-ray diffraction data. I would like to thank Annie Heroux and Alex Soares for their support during data collection. Tom Terwilliger assisted me in using the software package PHENIX, which he helped design, to solve the apo-structure of DpgC.

Finally, I would like to thank all members of my family especially my parents for emphasizing the importance of education in my life, and my wife for all her love and support during graduate school. I dedicate this thesis to my parents and my wife.

Table of Contents

Table of Figures

Chapter 2:

Chapter 4:

Chapter 5:

Abbreviations

4-OH-PheGly	4-Hydroxyphenylglycine
A domain	Adenylation domain
AmCoA	Amino coenzyme A
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
β-OH-Tyr	beta-hydroxy-tyrosine
βME	beta-mercapto ethanol
C domain	Condensation domain
CoA	Coenzyme A
Da	Dalton
DCM	Dichloromethane
DIPEA	Diisopropylethylamine
DpaCoA	Dihydroxyphenylacetyl coenzyme A
DPCK	Dephosphocoenzyme A kinase
Dpg	(S)-3,5-dihydroxyphenylglycine
EDTA	Ethylenediaminetetraaectic acid
EPR	Electron paramagnetic resonance
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
Hod	1H-3-Hydroxy-4-oxoquinaldine 2,4-dioxygenase
HPLC	High pressure liquid chromatography

MAD	Multi-wavelength anomalous dispersion
MeOH	Methanol
MIR	Multiple isomorphous replacement
NAD	Nicotinamide adenine dinucleotide
NMP	N-methyl pyrollidinone
NMR	Nuclear magnetic resonance
NRPS	Nonribosomal peptide synthetase
NSLS	National Synchrotron Light Source
Orf	Open reading frame
PanK	Pantetheine kinase
PCP domain	Peptidyl carrier protein domain
PKS	Polyketide synthase
PPAT	Phosphopantetheine adenylyl transferase
Qdo	1H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase
SAD	Single-wavelength anomalous dispersion
SeMet	Seleno-methionine
TE domain	Thioesterase domain
THF	Tetrahydrofuran
UV	Ultraviolet
VRE	Vancomycin resistant Enterococci

Chapter 1: Non-ribosomal peptides and the biosynthesis of non-proteinogenic amino acids through cofactor independent oxygenation.

Introduction

Vancomycin was originally isolated in the 1950s from soil samples found in Borneo.[1] Several members of this family of antibiotics from *Amycolatopsis* and *Streptomyces* strains have been isolated and structurally characterized.[2] Their complex molecular architecture has made this family of natural products of longstanding interest to both biosynthetic and synthetic chemistry efforts.[3,4,5]

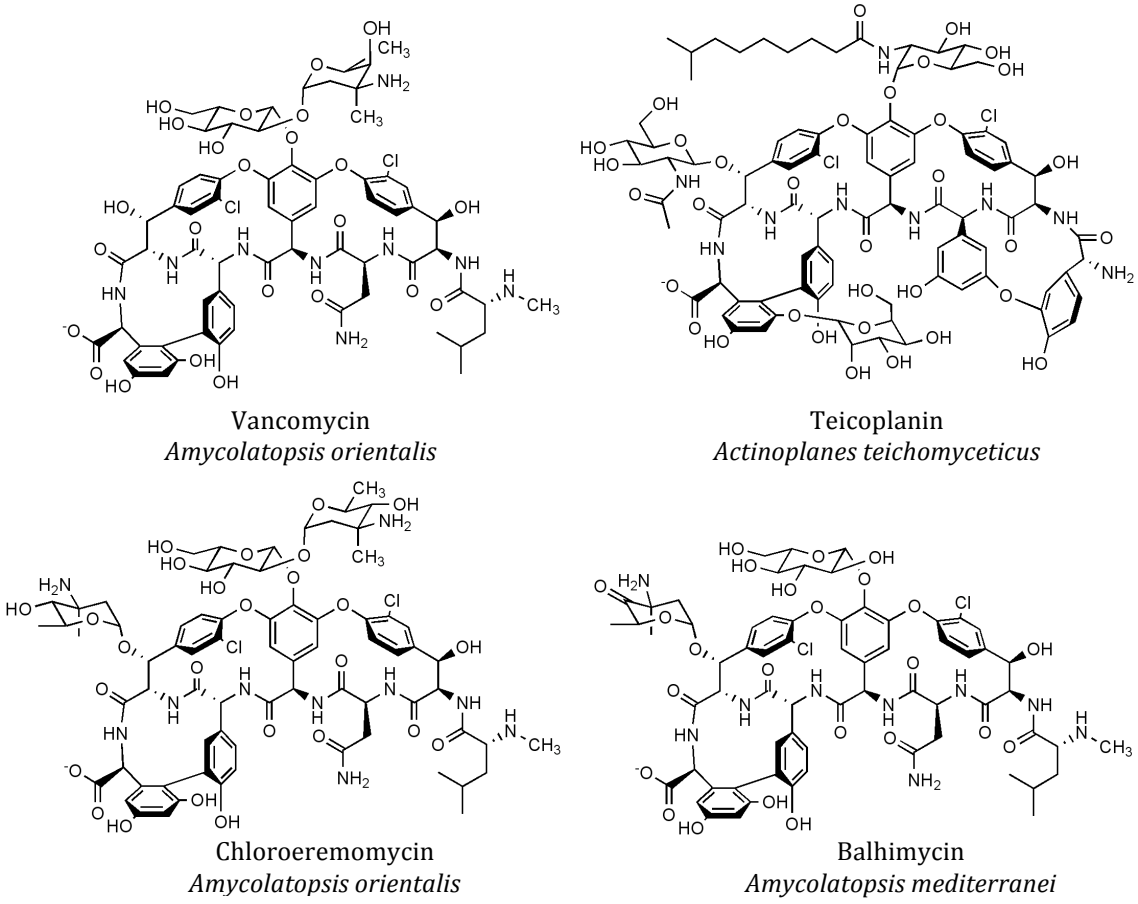


Figure 1: Vancomycin and teicoplanin family members, and their producing organism.

The vancomycin family of clinically important antibiotics are peptide based natural products commonly referred to as a "last resort" when treating life-threatening infections of gram-positive bacteria.[6] This family of antibiotics targets cell wall

biosynthesis through binding the D-Ala-D-Ala precursor of the growing peptidoglycan layer of the bacterial cell wall (Figure 2).[7] At its core vancomycin is a heptapeptide scaffold produced by a nonribosomal peptide synthetase (NRPS).[8] This scaffold undergoes extensive modification to give the mature natural product its potent biological activity. The key modifications to the peptide scaffold are the one aryl and two aryl ether cross-links that create the cup-like shape of vancomycin.[9] This rigidification allows

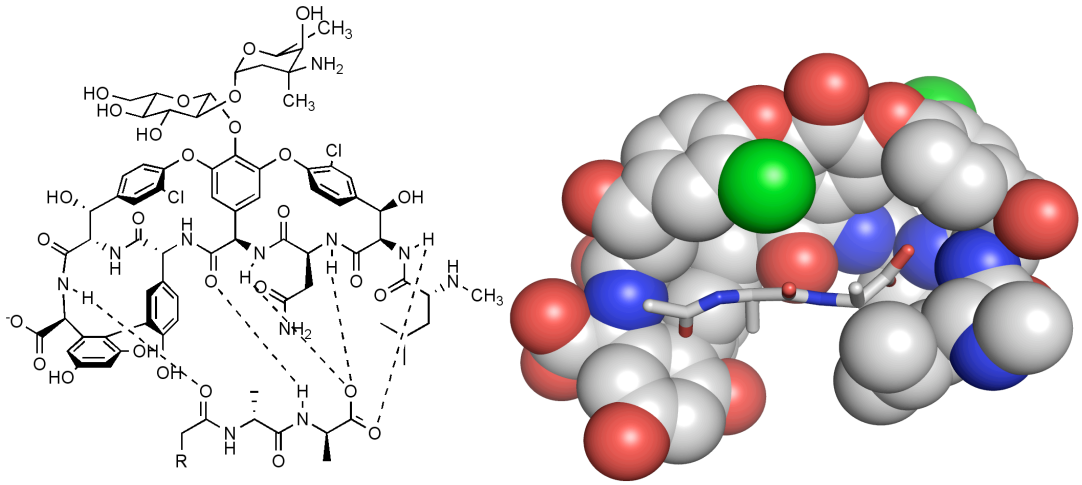


Figure 2: Vancomycin in complex with D-Ala-D-Ala (left), a space-filling model (right).

vancomycin to sequester the D-Ala-D-Ala precursor, thereby preventing bacterial cell wall biosynthesis. Vancomycin family members are also ornamented with sugar moieties at the 4-OH-PheGly$_4$ position, as well as the β-OH-Tyr$_6$ position in some family members (Figure 1).[5]

Most bacteria that produce small molecule antibiotics as chemical warfare agents in the battle for survival have a self-protection and immunity mechanism against their own antibiotics.[10] In the case of the vancomycin family producer organisms, self-protection comes from a mutation of the D-Ala-D-Ala cell wall precursor to a D-Ala-D-Lac moiety. It is thought that the source of resistance found in pathogenic *Enterococci* comes from the three genes found in the producing organisms that protect against

vancomycin.[11] Of the three resistance-conferring enzymes, VanH forms D-lactate from pyruvate, while DdlM couples D-lactate to D-alanine. The third enzyme, VanX, hydrolyzes D-Ala-D-Ala, making D-Ala-D-Lac a more readily abundant metabolite for cell wall biosynthesis.[12] The incorporation of a lactate moiety not only creates a loss of one hydrogen bond between vancomycin and the cell wall precursor; it also creates repulsion between a carbonyl in the vancomycin backbone and the ester oxygen in D-Ala-D-Lac (Figure 3). The rise of vancomycin-resistant *Enterococci* (VRE) has made the biosynthesis of this antibiotic family of particular interest. Producing vancomycin analogs holds great potential for overcoming resistance in VRE.[13] Through better understanding nature's method of synthesizing these complex natural products, a facile route toward analog synthesis could be developed.

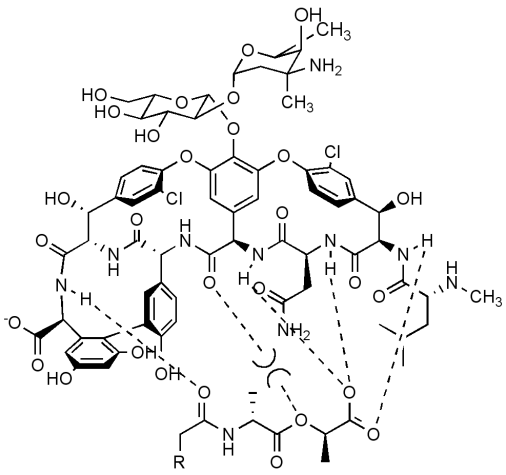


Figure 3: Vancomycin interacting with D-Ala-D-Lac

Nonribosomal peptide synthetases: Synthesis of the vancomycin heptapeptide

At the center of the vancomycin biosynthetic pathway are multidomain nonribosomal peptide synthetases (NRPS). These systems are organized into modules responsible for activation and incorporation of amino acid monomers (Figure 4).[14]

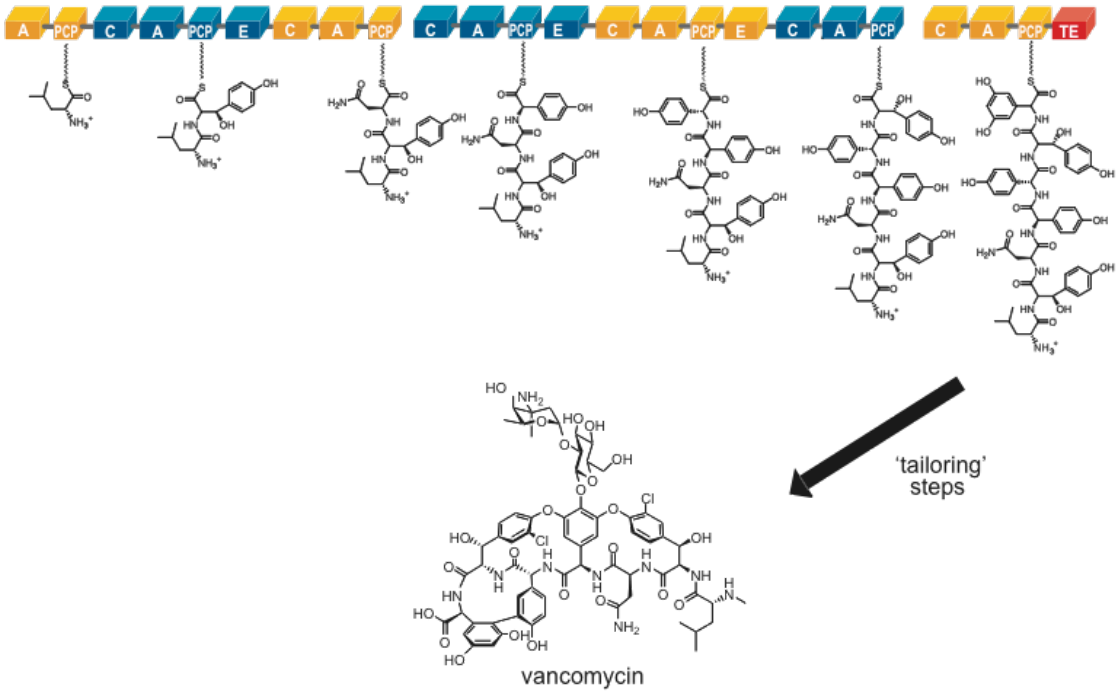


Figure 4: A schematic representation of the vancomycin non-ribosomal peptide synthetase (NRPS) assembly-line

Each module is made up of domains, individually folded proteins covalently attached to each other via linker regions. Each domain carries out a different aspect of constructing the vancomycin scaffold. Adenylation (A) domains activate specific amino acids by forming an aminoacyl-AMP, then transferring the aminoacyl group to a peptidyl carrier protein (PCP) equipped with a phosphopantetheinyl arm (Figure 5). The newly formed thioester bond then acts as an electrophile for the subsequent coupling reaction carried out by a condensation (C) domain. Seven A domains activate the distinct amino acid monomers that are coupled via amide bonds to form the heptapeptide scaffold. These A domains are in a co-linear sequence in the gene cluster corresponding to the order of

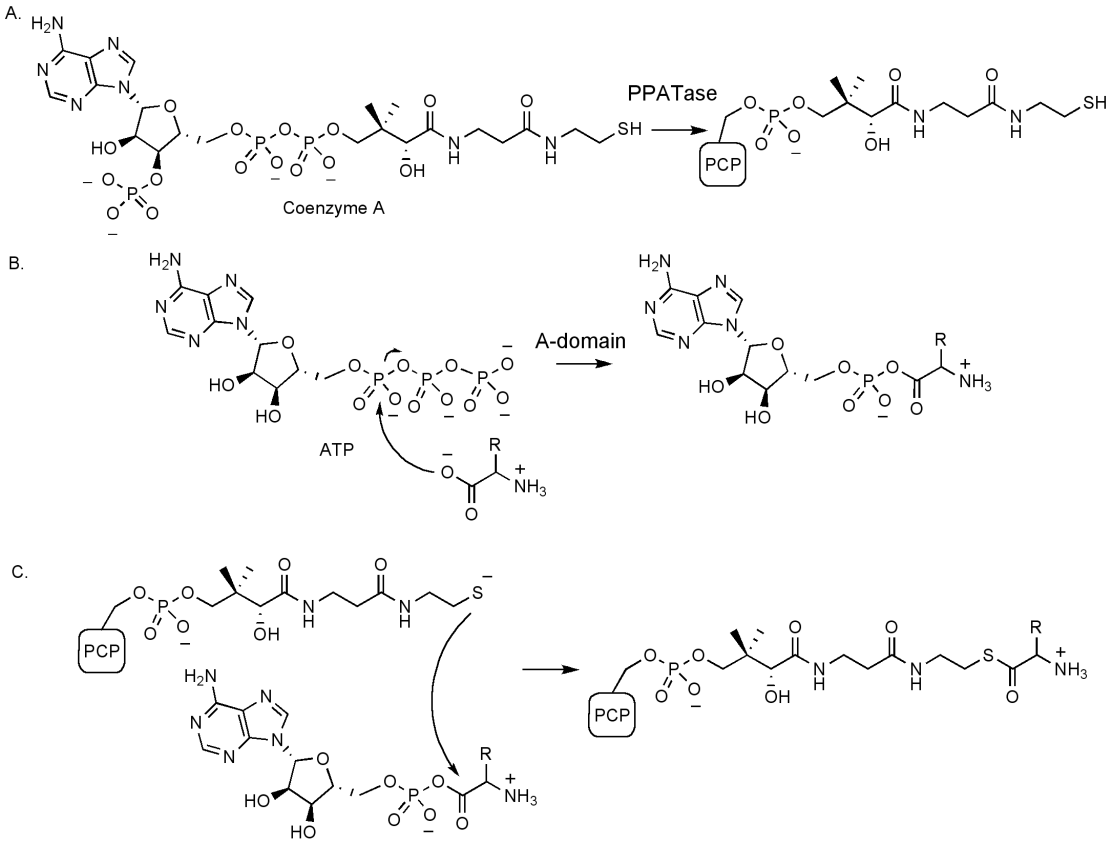


Figure 5: Activation and tethering of amino acids on the NRPS assembly-line; the chemistry of the adenylation (A) and peptidyl carrier (PCP) domains

amino acids in vancomycin. After each amino acid is loaded onto the synthetase, six C domains in the vancomycin pathway form the six peptide bonds.[15]

Not all NRPS systems have an even ratio between C domains and bonds formed by the assembly-line; one such case is discussed in Chapter 5 of this thesis.[16] Once a C domain makes a peptide bond, the growing peptide chain is passed down the assembly line to the following PCP domains. The final NRPS module contains a thioesterase (TE) domain, which cleaves the completed peptide scaffold from the NRPS assembly line.

Efforts towards engineering these systems have focused on exploiting the assembly-line nature of NRPS, and alteration of the glycosyl groups decorating vancomycin. Attaching a biphenyl group to the sugar moiety in the natural product

overcame resistance in vancomycin resistant enterococci (Figure 6).[17] The Kahne group demonstrated that this analog was not binding the D-Ala-D-Lac cell wall precursor, instead targeting bacterial transglycosylases.[18] These findings show that altering the vancomycin scaffold can lead to antibiotic function outside the typical mode of action for vancomycin. Study and manipulation of the biosynthetic pathway can potentially lead to overcoming resistance through unexpected means. The Liu group demonstrated that altered A domain specificity can be achieved through directed evolution.[19] By controlling which amino acid monomers A domains activate, this technique expands monomers available for incorporation by the NRPS system at various positions on the heptapeptide. By incorporating a variety of monomers at a selected position, biosynthesis of analogs could become quite facile.[20] This expansion of heptapeptide diversity makes the study of nonproteinogenic amino acids crucial to vancomycin analog biosynthesis.

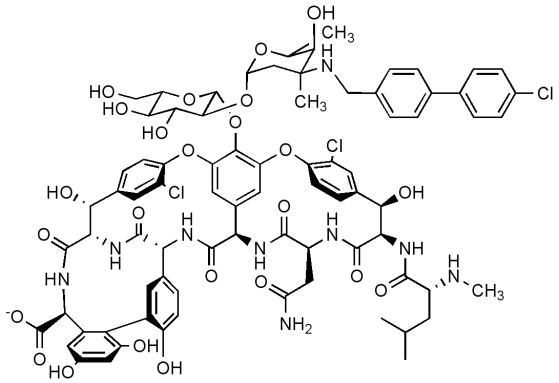


Figure 6: Chlorobiphenyl vancomycin; effective against VRE

NRPS systems incorporate over 300 amino acids to achieve a wide variety of peptide scaffolds, leading to a large structural diversity in peptide natural products. The biosynthesis of such building blocks is another interesting and complex aspect of vancomycin biosynthesis.

Biosynthesis of the nonproteinogenic amino acid (*S*)-3,5-dihydroxyphenylglycine, Dpg

The nonproteinogenic amino acid (*S*)-3,5-dihydroxyphenylglycine (Dpg) is incorporated at the seventh position in vancomycin family members and at the third and seventh positions in teicoplanin (Figure 1 and 7).[21] Unlike many nonproteinogenic amino acids found in natural products,[22] Dpg is not derived from the pool of 20 canonical amino acids. The biosynthesis of this amino acid is of particular importance due to the flexibility of the vancomycin biosynthetic pathway towards incorporating analogs into the seventh position. The Sussmuth group showed that supplementing a DpgA knockout mutant, which cannot produce Dpg, with various phenylacetic acid derivatives leads to the generation of novel glycopeptides through substitution at the seventh position.[23]

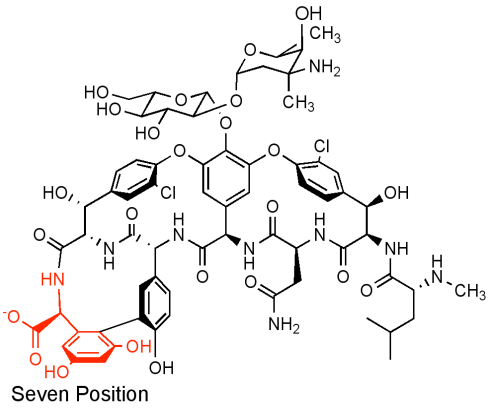


Figure 7: Vancomycin with Dpg, at the seventh position, highlighted

Since vancomycin analogs are known to have alternate antibiotic function and because of the plasticity at the seventh position, the pathway toward Dpg is of particular interest for many combinatorial biosynthesis efforts.

The biosynthetic gene clusters of the vancomycin family contain five enzymes involved in the biosynthesis of the amino acid Dpg. Elucidation of this pathway began with feeding experiments and culminated in biochemical characterization of all five enzymes. By supplementing fermentation of *Amycolatopsis orientalis* with (1,2-^{13}C)acetate, the Williams lab initially showed the source of the Dpg carbon skeleton is

derived from four acetate units.[24] This is in contrast to other nonproteinogenic amino acids in vancomycin and other non-ribosomal peptides, which are derived from

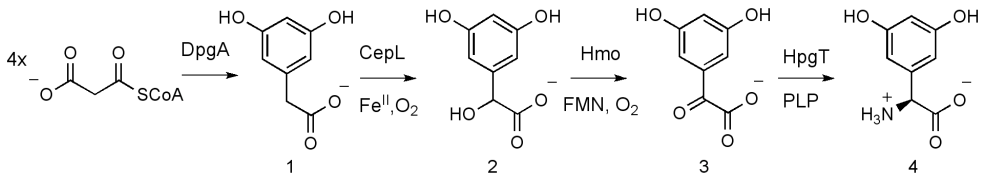


Figure 8: Dpg biosynthetic pathway proposed by the Williams group

proteinogenic amino acids like tyrosine. Additional feeding experiments were carried out in an effort to determine the intermediates of the Dpg pathway. It was shown that addition of ^{13}C labeled dihydroxyphenylacetic acid (**1**) and dihydroxymandelic acid (**2**) led to incorporation of these precursors into chloroeremomycin.[25] Based on this evidence and the putative assignment of open reading frames (ORF) in the gene cluster, the pathway shown in Figure 8 was proposed. Although this pathway does not resemble what we now know as the Dpg biosynthetic pathway, it is not totally unreasonable. Assuming that **2** and **3** are precursors to vancomycin, the ORFs selected to complete the biosynthesis of Dpg were logical selections. CepL, previously designated Orf20, is a heme dependent monooxygenase, which hydroxylates the benzyl position of L-Tyr to make β-hydroxytyrosine. In the Williams group proposal, CepL catalyzes a homologous reaction. However this transformation is unlikely as the actual substrate of CepL is tethered to a peptidyl carrier protein via a phosphopantetheine linker. (CepL is discussed in greater detail in the appendix of this thesis.) Unpublished data from the Williams group shows the FMN dependent oxidase Hmo, found in all vancomycin pathways, does in fact oxidize **2** to **3**. The final proposed step implicates the aminotransferase HpgT in

the amination of the glyoxylate; this step has been corroborated by subsequent research. A more likely explanation for the incorporation of ^{13}C labeled dihydroxyphenylacetic acid (**1**) is the action of a promiscuous CoA-ligase that couples **1** to CoA. This substrate would then follow the accepted pathway to Dpg. The incorporation of **2** may actually follow the proposed pathway, however this is not a naturally occurring pathway to Dpg. This work did identify Orf27, now called DpgA, as a chalcone synthase homolog, part of the Type III polyketide synthase (PKS) family, which couples four units of malonyl-CoA in the first step toward Dpg (Figure 9).

The Pelzer group confirmed that DpgA is in fact a Type III PKS responsible for the synthesis of a Dpg precursor.[26] Before exploration of the Dpg pathway, only one

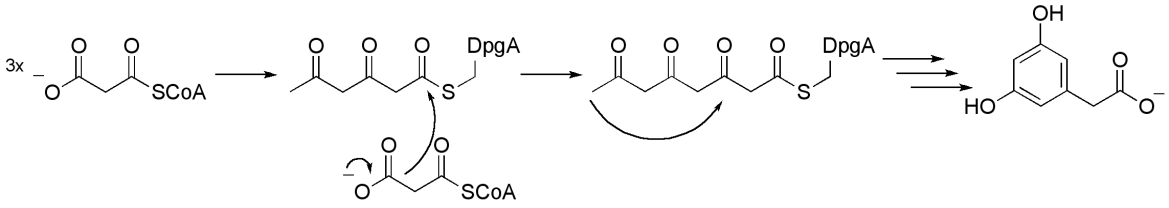


Figure 9: Proposed *in vivo* function of DpgA

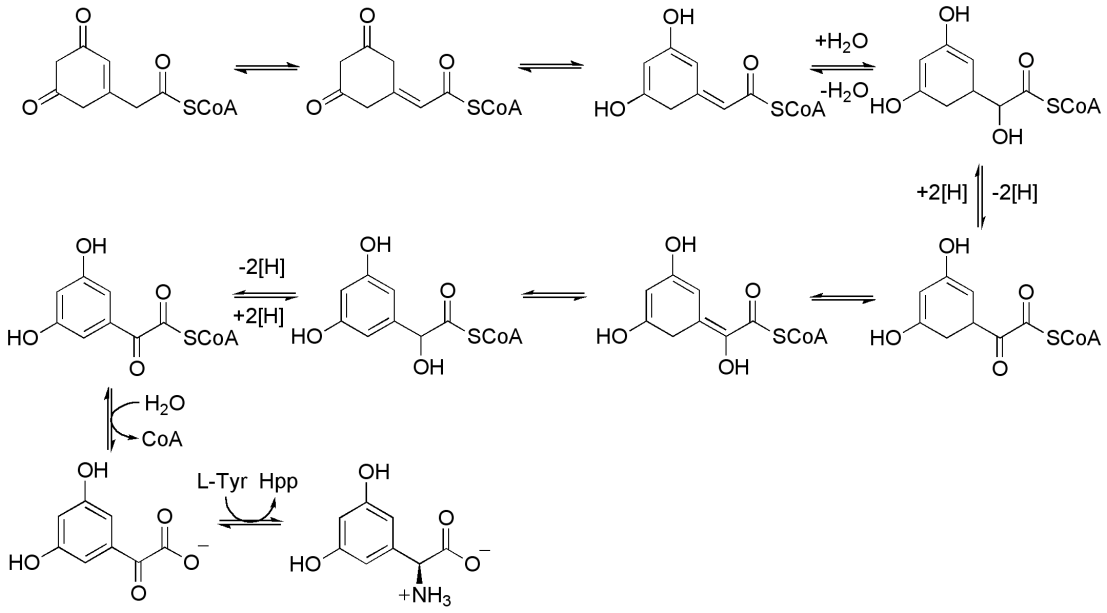


Figure 10: Dpg biosynthetic pathway proposed by the Pelzer group

Type III PKS had been indentified in prokaryotes.[27] The Pelzer group sought to characterize the Dpg pathway through genetic and biochemical studies. A deletion mutant of the *dpgA* gene in the producing organism, *A. mediterranei*, lacked the ability to produce the biologically active compound. Production of the antibiotic was restored to the mutant bacteria through addition of dihydroxyphenylacetic acid. Upon heterologous expression of DpgA in *Streptomyces lividans*, dihydroxyphenylacetic acid was produced. *In vitro* characterization of DpgA showed the production of dihydroxyphenylacetic acid exclusively from malonyl-CoA. Co-expression of DpgA, B, C and D in *Streptomyces lividans* led to production of dihydroxyphenylglyoxylate. Based on sequence homology the Pelzer group suggested that DpgB and D were involved in isomeration of the DpgA product while DpgC may possess dehydrogenase activity. A detailed biosynthetic pathway was proposed by the Pelzer group based on their characterization of DpgA and alignment studies of DpgB-D (Figure 10). The four enzymes, DpgA, DpgB, DpgC and DpgD, are translationally coupled with *dpgA* the beginning of the operon. DpgA-D have sequence similarity to enzymes that use coenzyme A derivatives as substrates, leading to the conclusion that the pathway is carried out on CoA derivatives as opposed to free acids. Therefore, it is interesting that the Pelzer group concluded the product of DpgA to be a free acid. The Pelzer group also analyzed a knockout of the aminotransferase HpgT. The *HpgT* knockout did not produce the antibiotically active compound. The pathway was only reconstituted through addition of both 4-hydroxyphenylglycine (Hpg) and 3,5-dihydroxyphenylglycine (Dpg), showing the use of HpgT in both amino acid pathways. This work offered insight into the function of DpgA. However biochemical analysis of the pathway was incomplete, leaving the proposed pathway open to interpretation.

The Walsh lab described the activity of the enzymes DpgA-D through heterologous expression in *E. coli* and *in vitro* biochemical characterization.[28] The

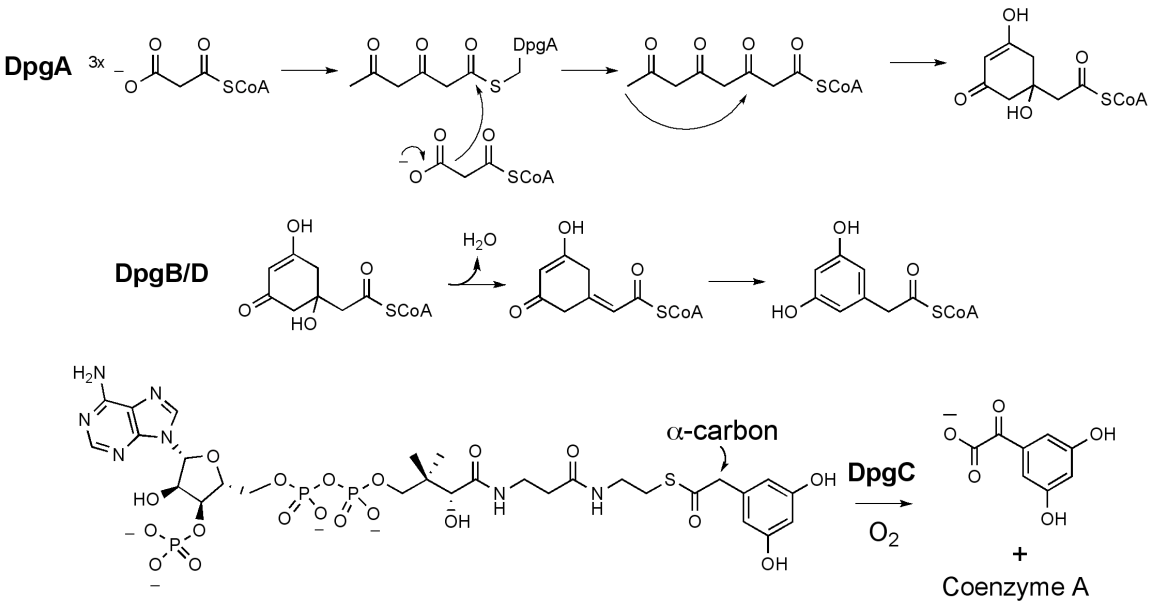


Figure 11: Reactions catalyzed by the enzymes DpgA, DpgB/D and DpgC proposed by the Walsh group

function of DpgA was properly assigned through incubation with malonyl-CoA, which resulted in formation of dihydroxyphenylacetyl-CoA (Dpa-CoA) as detected by HPLC analysis. The rate of reaction of DpgA was very slow and required overnight incubations to detect any product. Upon addition of DpgB, a 17-fold increase in rate was detected, although DpgB alone has no detectable reactivity towards malonyl-CoA. When DpgD was added to the DpgA/B mixture, a further 2-fold increase was observed in the production of Dpa-CoA. A DpgA/D mixture did not result in any rate increase over the DpgA rate. Bioinformatic analysis of DpgB and DpgD showed that they had high similarities to the crotonase superfamily. Hydratase activity, common in the crotonase

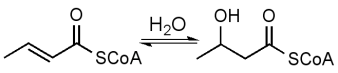


Figure 12: Enoyl-CoA hydratase (crotonase) activity

superfamily (Figure 12), was detected when DpgB and DpgD were incubated with crotonyl- and β-methylcrotonyl-CoA. Due to the rate enhancement DpgB and DpgD demonstrate, they could possibly form a complex with DpgA. Details on how these crotonase family enzymes accelerate the rate of Dpa-CoA production are yet to be resolved.

This initial publication on the Dpg pathway also included some characterization of the enzyme DpgC. The unknown function of this unusual enzyme was largely responsible for the limited understanding of Dpg biosynthesis. The Walsh group demonstrated the enzyme DpgC oxidizes the α–C methylene to a ketone and cleaves the thioester of Dpa-CoA to release dihydroxyphenylglyoxylate and CoA (Figure 11). Three separate enzymes typically carry out this transformation: flavoprotein desaturase, crotonyl-type enoyl thioester hydratase and an alcohol dehydrogenase. These initial results warranted further study of DpgC by the Walsh group, discussed in the next section.

The cofactor-independent dioxygenase DpgC

As discussed above, DpgC performs a four-electron oxidation and cleaves a thioester bond (Figure 11). Detailed biochemical characterization of DpgC demonstrated it catalyzes this unique oxidation chemistry using only molecular oxygen and the substrate. The absence of any cofactor or metal puts DpgC in a very small subset of oxygenases, which makes it a fascinating target for study by mechanistic enzymology.

Bioinformatic analysis of DpgC indicated the C-terminal two-thirds of the enzyme is homologous to crotonase superfamily members. It is interesting that the

pathway to Dpg contains enzymes involved in cyclization and in oxygenation that are all members of the same enzyme superfamily. Additional examples of oxygenases and cyclases from the same biosynthetic pathway arising from similar protein scaffolds are discussed later in this chapter.

Labeling studies with ^{18}O established that DpgC incorporates both oxygen atoms from molecular oxygen into the glyoxylate product.[29] Enzyme activity was shown to be dependent only on the presence of O_2 and no turnover was observed under an argon atmosphere or in degassed buffers. DpgC functions in the presence of large concentrations of ethylenediaminetetraaectic acid (EDTA) and atomic absorption analysis failed to reveal any stoichiometric bound metals. UV/visible absorbance measurements as well as mass spectrometry data indicated an absence of any cofactor such as flavin, pterin or iron. Deuterium exchange assays showed that the α–C hydrogens of Dpa-CoA exchange with solvent when incubated with DpgC under anaerobic conditions (Figure 13). Crotonase enzymes process CoA-thioester substrates through enolate stabilization, this fact leads to the proposal that DpgC binds a Dpa-CoA enolate. This electron rich intermediate would then react with O_2 to generate the product. These biochemical data raise the crucial question: How does DpgC activate molecular oxygen for reaction with organic substrates? A crystal structure of DpgC bond to a substrate mimic would answer how it performs this chemistry without a metal or cofactor. Understanding the way organic

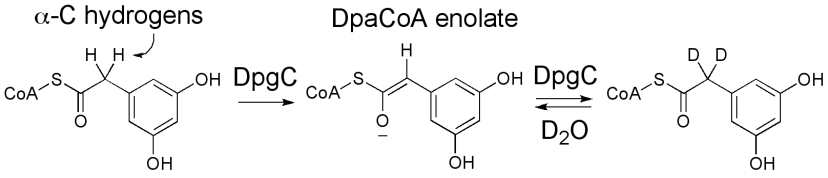


Figure 13: Deuterium exchange assay with DpgC and substrate

cofactors assist oxygenases is instructive as to how DpgC might function. Biochemical work on other cofactor and metal independent oxygenases also provides some insight into how such a unique enzyme carries out this chemistry.

II. Activation of molecular oxygen in nature

The role of cofactors in enzyme catalyzed oxidations

Chemical transformations involving molecular oxygen and organic substrates are complex processes. Molecular oxygen is in the triplet state, making it inert toward direct reaction with the vast majority of organic molecules, which are in the singlet state.[30] A reaction between two molecules in different spin states violates the conservation of angular momentum. Enzymes use a variety of techniques to overcome this barrier. Metalloenzymes use tightly bound transition metals, typically iron or copper, to catalyze a variety of transformations. Interesting examples of metalloenzyme chemistry in

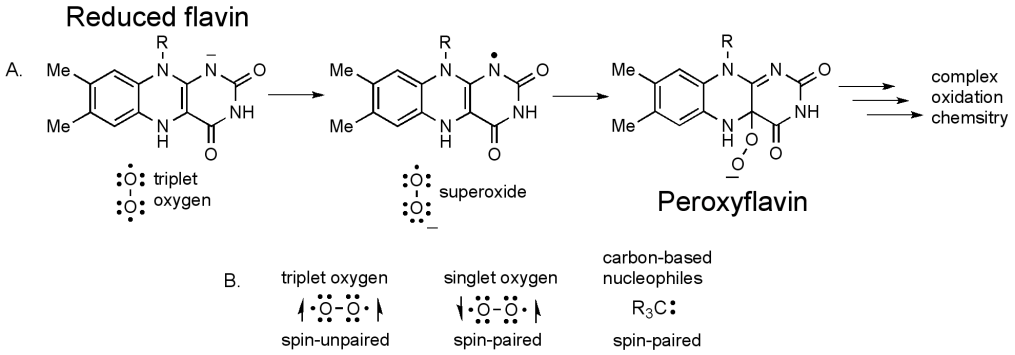


Figure 14: Reduction of molecular oxygen to superoxide by reduced flavin, spin states of oxygen and a carbanion

vancomycin biosynthesis are discussed in the appendix of this thesis. Enzymes also use the organic cofactor flavin in a wide array of chemical reactions. Applications of flavoenzymes range from histone modification to light emission.[31] In its reduced form

flavin represents an electron rich species capable of reducing molecular oxygen through single electron transfer (Figure 14). Following the reduction of molecular oxygen to superoxide, a common intermediate is a peroxyanion. This species has been identified in studies of flavin-dependent monooxygenases, enzymes that insert one atom of molecular oxygen while reducing the other oxygen atom to H_2O.

Baeyer-Villiger monooxygenases use flavin, NADPH, and molecular oxygen to catalyze the insertion of an oxygen atom into a carbon-carbon bond (Figure 13). This class of enzymes is of particular biotechnological interest due to the wide use of Baeyer-Villiger reactions in synthetic organic chemistry. Kinetic and spectroscopic studies have shown the enzymatic mechanism is conceptually identical to the nonenzymatic mechanism.[32] The key feature of the enzymatic mechanism is the reaction of reduced flavin with molecular oxygen to produce a stable flavin-peroxide intermediate.[33] The Mattevi group solved the crystal structure of a Baeyer Villiger monooxygenase, phenyl acetone moonoxygenase (PAMO).[34] PAMO showed the highest structural similarity to the flavoprotein class of disulfide oxidoreductases, which are comprised of two domains: a flavin binding domain, and an NADPH binding domain. PAMO makes van der Waals contacts through several aromatic amino acids (Trp55,

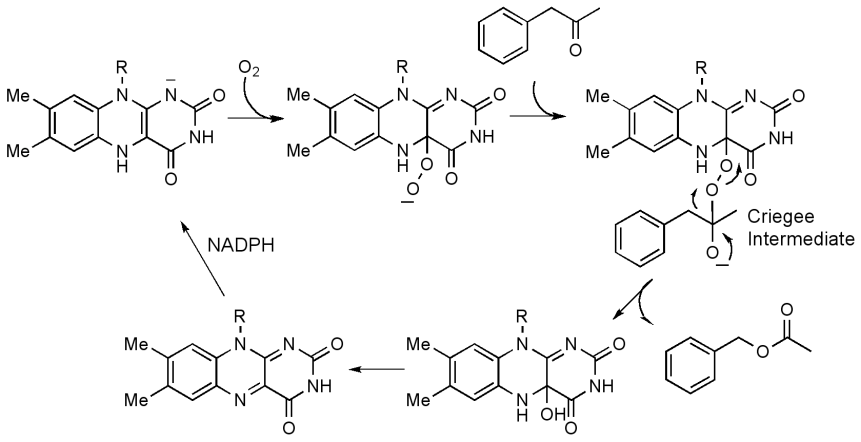


Figure 15: Catalytic cycle of the Baeyer-Villiger monooxygenase PAMO

Tyr60 and Tyr72) on the *si* face of the bound flavin. On the *re* face of flavin, Arg337 forms a π-stacking interaction with the cofactor. Arg337 is crucial to the catalytic activity of PAMO, as evidenced by the Arg337Ala mutant that lacks activity. It is believed that once flavin is reduced, Arg337 stabilizes the negatively charged deprotonated intermediate, and then stabilizes the peroxanion intermediate. Arg337 may also be involved in the Criegee intermediate that leads to product formation.[34]

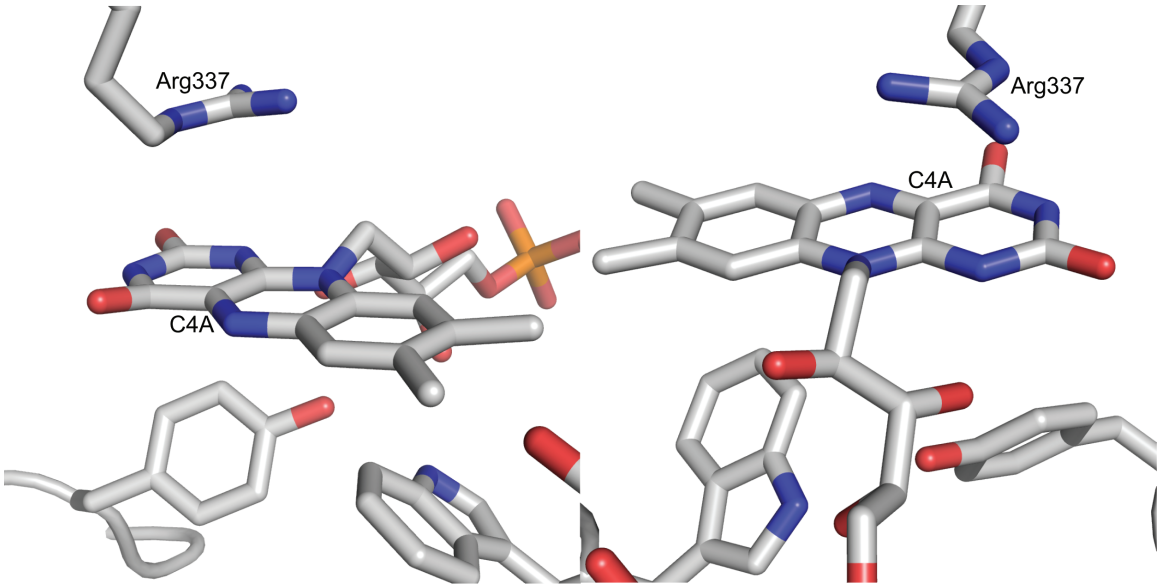


Figure 16: The active site of PAMO. Arg337 over the C4A of flavin, ideally situated to deprotonate the substrate and stabilize the flavinperoxy intermediate. The opposite face of the flavin makes van der Waals contacts with aromatic residues.

Glucose oxidase (GO) is a flavin dependent enzyme that catalyzes a hydride transfer from the anomeric C-H bond of glucose to FAD and the oxidation of $FADH^-$ by O_2. The rate-limiting step of this reaction is the single electron transfer from the negatively charged cofactor to molecular oxygen to from a caged radical pair. A crystal structure of GO revealed a histidine residue situated above the C4a-N5 position of the flavin molecule.[35] Kinetic studies of His516 mutant constructs showed this residue to be chiefly responsible for catalyzing the reaction between $FADH^-$ and O_2. In fact, a general

feature of flavoprotein oxidases is a correlation between the rate of O_2 activation by $FADH^-$ and a positively charged residue in the active site.

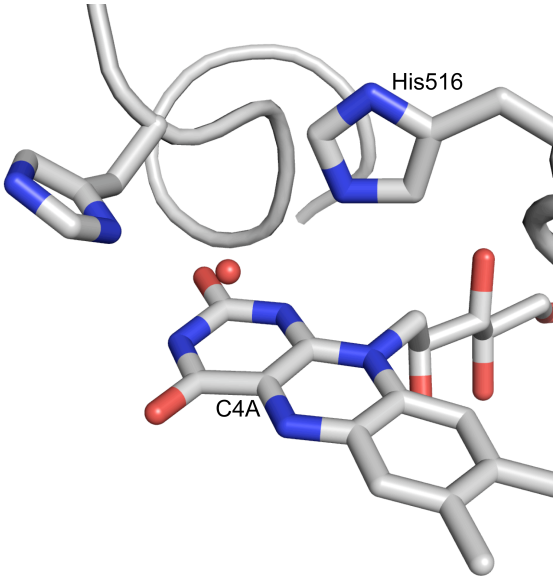


Figure 17: The active site of glucose oxidase. The location of the water molecule, red sphere is believed to mimic the peroxy intermediate

The critical role arginine and histidine residues play in deprotonation and stabilization of flavin cofactors is instructive to the way in which cofactor-independent oxygenases carry out anion stabilization. The path of negative charge accumulation followed by single electron transfer to molecular oxygen is thought to be shared by flavoenzymes and cofactor-independent oxygenases.

Cofactor and metal independent oxygenases

Oxygenases directly incorporate either one or two atoms of oxygen from molecular oxygen into their substrates. These enzymes are involved in biological processes ranging from degradation of natural and xenobiotic compounds to secondary metabolite tailoring.[36] Because the chemistry of these enzymes is regioselective and stereospecific, they are attractive targets for adaptation to a wide range of reactions. The vast majority of characterized oxygenases use transition metals or organic cofactors to activate triplet molecular oxygen for diverse oxidation chemistry. However, there is a small class of enzymes that perform oxidation chemistry without the use of a metal or

cofactor. The study of these enzymes leads to the question of how molecular oxygen is activated without prosthetic groups such as iron, copper or flavin cofactors.

Cofactor-independent monooxygenases in polyketide biosynthesis

The synthesis of polyketide natural products draws several analogies to NRPS biosynthesis because of the assembly-line nature of the polyketide synthase (PKS) systems.[37] For this reason there has been a great deal of work surrounding the manipulation of such systems towards combinatorial biosynthesis. Tailoring steps in polyketide biosynthesis are the subject of engineering efforts due to the potential for analog development. Cofactor-independent oxygenases identified from polyketide tailoring steps provide insight towards enzymatic oxygen activation. In 1993, Shen and Hutchinson isolated tetracenomycin F1 monooxygenase (TcmH) from *Streptomyces glaucescens*.[38] This small, 12.6 kDa enzyme catalyzes the oxidation of naphthacenone tetracenomycin F1 at position C-5 to form 5,12-napthacenequinone tetracenomycin D3 (Figure 18). It was determined that TcmH simply requires molecular oxygen and the substrate to carry out this reaction. One atom of molecular oxygen is incorporated into the product while the other is reduced to water. Based on initial biochemical characterization of the enzyme, it was thought that a histidine residue was involved in deprotonation of the substrate and stabilization of an

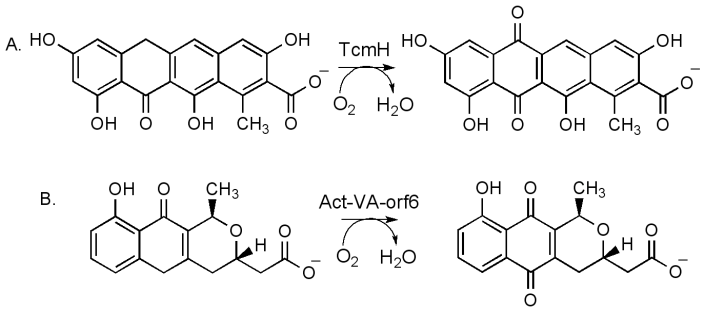


Figure 18: Chemistry of the cofactor-independent monooxygenases (A) TcmH and (B) ActVa-orf6

anionic intermediate. Greater detail on the mechanism of this monooxygenase was provided by the structure of a homolog, ActVA-Orf6, from *Streptomyces coelicolor*.[39]

The monooxygenase ActVA-Orf6 catalyzes the oxidation of 6-deoxydihydrokalafungin (6-DDHK) to dihydrokalafungin (DHK), and is 39% identical to TcmH. ActVA-Orf6 also catalyzes the same oxidation as TcmH on tetracenomycin F1 (Figure 18). As with TcmH, a histidine residue was assumed to play a major role in the catalysis of ActVA-Orf6. The mutation of a conserved histidine to a His52Gln construct showed total loss of activity in ActVA-Orf6. This initial biochemical data furthered the theory of a histidine residue playing the role of substrate deprotonation and anion stabilization.

The X-ray crystal structures of ActVA-Orf6 were solved unbound as well as bound to four substrate and product analogs. The overall fold of ActVA-Orf6 resembles the ferredoxin family of enzymes. This enzyme is the first example of an oxygenase in the ferredoxin family, representing another functionality for this diverse group of enzymes. The structure showed that the highly conserved His52 residue is not involved in catalysis; instead it makes contacts at the dimer interface. The formation of a dimer in ActVA-Orf6 is especially important as a C-terminal β-strand contributes to the β-sheet of the other monomer, completing the active site. Although structural analysis of His52 ruled out a catalytic role similar to a flavin dependent enzyme, analysis of the active site bound to substrate and product analogs lead to a mechanistic proposal similar to flavin chemistry.

The active site of ActVA-Orf6 did not display any electron density consistent with an organic cofactor, metal or metal-binding site, supporting the claim of ActVA-

Orf6 being a cofactor-independent monooxygenase. Residues identified in the active site as being important for catalysis were Tyr51, Asn62, Trp66 and Tyr72. Trp66 and Tyr72

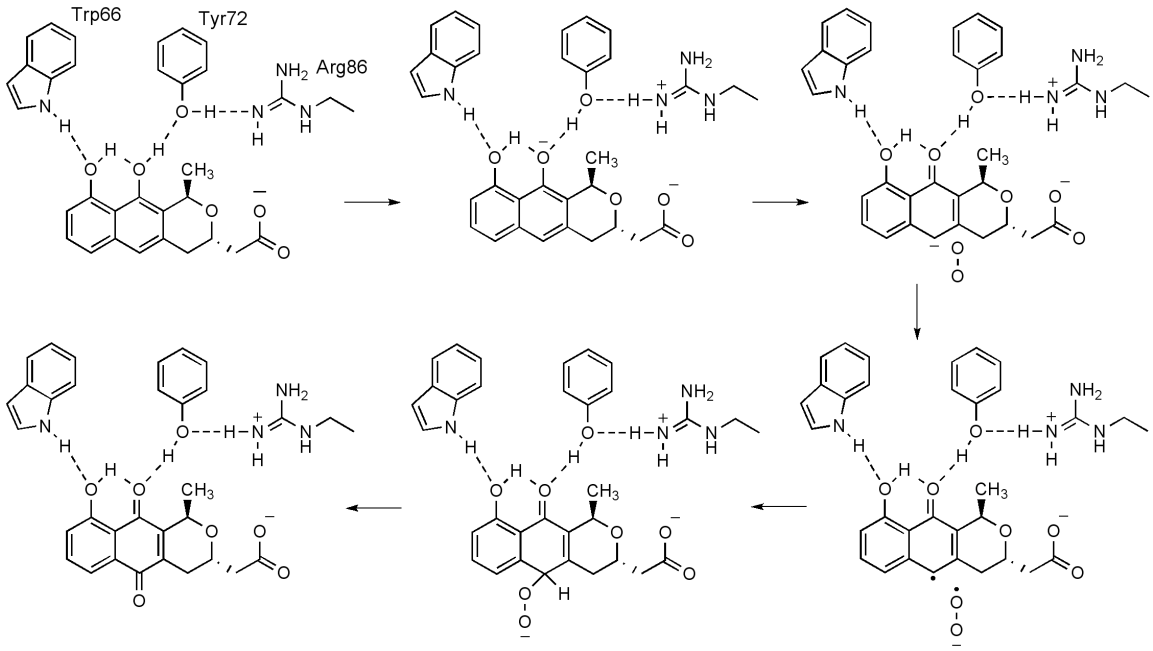


Figure 19: Proposed mechanism of ActVA-orf6

form hydrogen bonds with the hydroxyl groups at C11 and C13, respectively. These interactions are crucial to orienting the substrate in the active site, and for creating a buildup of negative charge on the substrate. Asn62 and Tyr61 orient a water molecule that facilites deprotonation at the site of oxidation. Sciara *et al*. propose that after substrate deprotonation, molecular oxygen is reduced by the substrate and then forms a peroxo intermediate stabilized by Asn62 (Figure 19).[39] This work provides some insight into substrate binding of a cofactor-independent oxygenase. However, the small molecules bound in the crystallographic studies differ from the substrate and may not give an accurate picture of the active site geometry upon substrate binding. The relaxed specificity of this class of cofactor-independent monooxygenases is evident in the

structural data of the active site. Very little concerning oxygen binding was gained from this work. With regard to how these enzymes bind and activate molecular oxygen very little was known at this point. An interesting aspect of this work is the identification of a ferredoxin fold for these enzymes, which was unexpected since sequence identity is low, less than 15% over fewer than 90 structurally equivalent amino acids. It had originally been proposed that TcmH and ActVA-Orf6 could represent a novel protein fold. The identification of these enzymes in the ferredoxin family shows the power of divergent evolution to develop broad functionality in an enzyme family.

Further examples of cofactor-independent oxygenases also come from polyketide biosynthesis, and are members of a different protein fold. Beinker *et al.* solved the structures of SnoaL2 and AclR, hydroxylases found in the nogalamycin and aklavinone biosynthetic pathways, respectively.[40] These hydroxylases of about 140 amino acids in size show no sequence similarity to flavin or metal-dependent hydroxylases. However, they do show approximately 25% sequence identity to polyketide cyclases. The closest homologues to SnoaL2 and AclR are the polyketide cyclases SnoaL and AknH found in the corresponding biosynthetic pathways. In the case of actinorhodin biosynthesis, a polyketide cyclase and cofactor-independent oxygenase was present both having the same protein fold. A strikingly similar case evolved in the nogalamycin and aklavinone biosynthetic pathways. A single protein scaffold gave rise to a cyclase and cofactor-independent oxygenase. Each case represents a different protein fold that has given rise to this functionality.

This evolution of function shows that a variety of protein scaffolds can bind and activate a substrate for cofactor-independent oxygenation. When such a scaffold is

present in a biosynthetic pathway, oxygenase function can evolve when required by a particular organism. It is unclear if the presence of cyclase and oxygenase from the same fold in a particular biosynthetic pathway has any significance. Do these two chemical functionalities have a correlation that makes them closely related from an evolutionary perspective?

Cofactor-independent dioxygenases in heteroaromatic degradation pathways

1*H*-3-Hydroxy-4-oxoquinaldine 2,4-dioxygenase (Hod) from *Arthrobacter ilicis* and 1*H*-3-hydroxy-4-oxoquinoline 2,4-dioxygenase (Qdo) from *Pseudomonas putida* are part of the anthranilate pathway of quinaldine and 4-quinoline degradation, respectively.[36] These enzymes open hetercyclic ring systems through the insertion of dioxygen at the C2 and C4 position, which leads to the cleavage of two carbon-carbon bonds and the release of carbon monoxide (Figure 20). Biochemical and spectroscopic characterization of Hod and Qdo have shown that neither contains a metal or organic cofactor.[41] Sequence analysis of these ring-cleaving dioxygenases shows no similarity to known oxygenases. Hod and Qdo are members of the α/β hydrolase superfamily of proteins. Along with luciferase, they are the only oxygenases in this catalytically diverse group of enzymes. The α/β hydrolase family contains a highly conserved catalytic triad found at a "nucleophilic elbow." This triad contains a nucleophile (Ser, Asp or Cys) an acidic residue (Asp or Glu) and an absolutely conserved

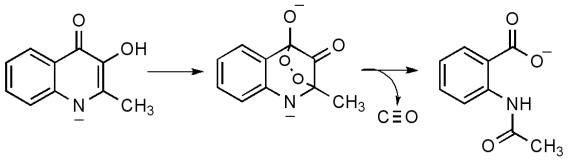


Figure 20: Chemistry of the cofactor-independent dioxygenase Hod

histidine residue. Hod and Qdo both have a histidine that aligns with the triad histidine of the α/β hydrolases and a serine as a potential nucleophile. These residues are found within a motif corresponding to the "nucleophilic elbow." The conserved histidine residue, His244 of Qdo, was found to be essential for catalysis, while the proposed conserved serine residue was not. An acidic residue that would complete the catalytic triad was never identified. This finding, in addition to the active serine mutant, suggests that Qdo and Hod do not contain the "canonical" α/β hydrolase catalytic triad.

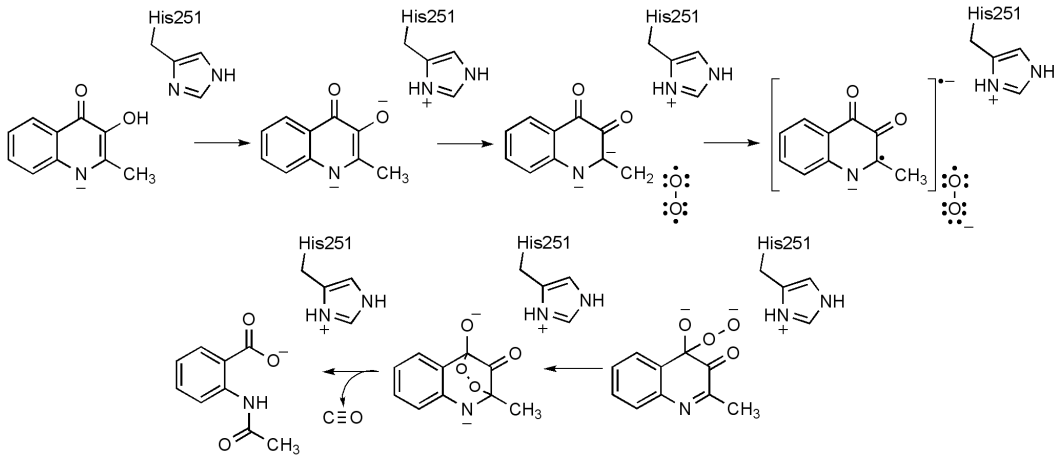


Figure 21: Mechanism of Hod as proposed by the Fetzner group

The Fetzner lab performed steady-state kinetic analysis of Hod to better understand the function of cofactor-independent dioxygenases.[42] They propose that His251, found in the conserved "nucleophilic elbow" position of α/β hydrolases, acts as a general base which deprotonates the substrate and stabilizes a dianion intermediate. Through electron paramagnetic resonance (EPR) studies, single electron oxidation of this high-energy intermediate was observed, which suggest the reduction of molecular oxygen to superoxide. This kinetic data also demonstrated that Hod binds the organic substrate first then molecular oxygen. Sequential substrate binding can be explained by a

conformational change taking place upon binding of the first substrate, which allows for binding of the second substrate. Conformation changes in flavin-dependent enzymes are thought to be required for reaction of reduced flavin with dioxygen. With the crucial role of a conserved histidine identified by this kinetic work, comparisons to flavin enzymes are appropriate. Based on this work, the Fetzner lab proposed a mechanism with a caged radical pair, even though caged radical pairs are undetectable by conventional EPR (Figure 19). In addition to activating the organic substrate and binding molecular oxygen, the function of Hod is also thought to provide an environment suitable for electron transfer. Structural characterization of a cofactor-independent dioxygenase would shed light on the nature of such an environment.

Crystallographic studies of the cofactor-independent dioxygenase DpgC

Questions of substrate recognition, binding of molecular oxygen and activation of molecular oxygen in cofactor-independent oxygenases remain largely unanswered. It is generally assumed the substrate plays a role in oxygen activation, yet how this is achieved was largely a mystery until recently. Our work on the crystal structure of DpgC bound to an isosteric substrate mimic reveals a great deal about oxygen binding and substrate activation. The active site environment binds molecular oxygen and is also suitable for single electron transfer. Our efforts represent the first crystallographic structure of a cofactor-independent dioxygenase. Not only does the study of this enzyme provide information on oxygen activation, it also teaches us about the synthesis of a crucial amino acid monomer in the powerful antibiotic vancomycin. The amino acid Dpg

is incorporated into a position in vancomycin amenable to substitution, which makes detailed knowledge of Dpg biosynthesis useful in the production of vancomycin analogs.

Chapter summaries:

Chapter 1: Introduction: Non-ribosomal peptides and the biosynthesis of non-proteinogenic amino acids through unique oxidation chemistry.

Chapter 2: Solving the apo-structure of DpgC: A cofactor independent dioxygenase in the vancomycin biosynthetic pathway. Crystallography behind solving the structure of DpgC and the chemistry learned from the apo-structure.

Chapter 3: Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis: Crystal structure of DpgC bound to an isosteric substrate mimic. Substrate recognition and catalysis identification based on the model of substrate mimic bound to DpgC.

Chapter 4: Probing the oxygen-binding pocket of DpgC. Biochemical and crystallographic studies of the DpgC oxygen-binding pocket. Proposal of two possible paths for molecular oxygen from the initial oxygen-binding pocket in the DpgC active site.

Chapter 5: Biochemical characterization and efforts toward structural characterization of a non-ribosomal peptide synthetase module from *Thermobifida*

***fusca*.** Discussion of the fuscachelin pathway, an NRPS pathway that produces a natural product recently elucidated in the Bruner lab. The NRPS module FscI was biochemically characterized for adenylation domain specificity and an effort towards structurally characterizing the multi-domain module was undertook through X-ray crystallography.

References

1 McCormick, M. H. et al., Vancomycin, a new antibiotic. I. Chemical and biologic properties. *Antibiot Annu* **3**, 606 (1955).

2 Sheldrick, G. M. et al., Structure of vancomycin and its complex with acetyl-D-alanyl-D-alanine. *Nature* **271** (5642), 223 (1978).

3 Pelzer, S. et al., Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908. *Antimicrob Agents Chemother* **43** (7), 1565 (1999); Pootoolal, J. et al., Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from Streptomyces toyocaensis NRRL15009. *Proc Natl Acad Sci U S A* **99** (13), 8962 (2002).

4 van Wageningen, A. M. et al., Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem Biol* **5** (3), 155 (1998).

5 Nicolaou, K. C., Boddy, C. N., Brase, S., and Winssinger, N., Chemistry, Biology, and Medicine of the Glycopeptide Antibiotics. *Angew Chem Int Ed Engl* **38** (15), 2096 (1999).

6 Anstead, G. M., Quinones-Nazario, G., and Lewis, J. S., 2nd, Treatment of infections caused by resistant Staphylococcus aureus. *Methods Mol Biol* **391**, 227 (2007).

7 Williams, D. H., Williamson, M. P., Butcher, D. W., and Hammond, S. J., Detailed Binding-Sites of the Antibiotics Vancomycin and Ristocetin-a - Determination of Intermolecular Distances in Antibiotic Substrate Complexes by

Use of the Time-Dependent Noe. *Journal of the American Chemical Society* **105** (5), 1332 (1983).

8 Marahiel, M. A., Stachelhaus, T., and Mootz, H. D., Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev* **97** (7), 2651 (1997).

9 Bischoff, D. et al., The Biosynthesis of Vancomycin-Type Glycopeptide Antibiotics-New Insights into the Cyclization Steps This work was supported by the Deutsche Forschungsgemeinschaft (SFB 323). We thank M. Schierle, Dr. S. Stevanovic and Prof. H.-G. Rammensee for help with Edman degradation and J. Turner, Prof. B. List and Prof. D. Boger (La Jolla, USA) for discussions on the work. *Angew Chem Int Ed Engl* **40** (9), 1693 (2001).

10 Davies, J., Inactivation of antibiotics and the dissemination of resistance genes. *Science* **264** (5157), 375 (1994).

11 Sosio, M., Bianchi, A., Bossi, E., and Donadio, S., Teicoplanin biosynthesis genes in Actinoplanes teichomyceticus. *Antonie Van Leeuwenhoek* **78** (3-4), 379 (2000).

12 Marshall, C. G. and Wright, G. D., DdlN from vancomycin-producing Amycolatopsis orientalis C329.2 is a VanA homologue with D-alanyl-D-lactate ligase activity. *J Bacteriol* **180** (21), 5792 (1998); Marshall, C. G., Lessard, I. A., Park, I., and Wright, G. D., Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. *Antimicrob Agents Chemother* **42** (9), 2215 (1998).

13 Fu, X., Albermann, C., Zhang, C., and Thorson, J. S., Diversifying vancomycin via chemoenzymatic strategies. *Org Lett* **7** (8), 1513 (2005).

14 von Dohren, H., Keller, U., Vater, J., and Zocher, R., Multifunctional Peptide Synthetases. *Chem Rev* **97** (7), 2675 (1997).

15 Hubbard, B. K. and Walsh, C. T., Vancomycin assembly: nature's way. *Angew Chem Int Ed Engl* **42** (7), 730 (2003).

16 Dimise, E. J., Widboom, P. F., and Bruner, S. D., Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile Thermobifida fusca. *Proc Natl Acad Sci U S A* **105** (40), 15311 (2008).

17 Biavasco, F. et al., In vitro antibacterial activity of LY333328, a new semisynthetic glycopeptide. *Antimicrob Agents Chemother* **41** (10), 2165 (1997).

18 Chen, L. et al., Vancomycin analogues active against vanA-resistant strains inhibit bacterial transglycosylase without binding substrate. *Proc Natl Acad Sci U S A* **100** (10), 5658 (2003).

19 Fischbach, M. A. et al., Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc Natl Acad Sci U S A* **104** (29), 11951 (2007).

20 Hutchinson, C. R. and McDaniel, R., Combinatorial biosynthesis in microorganisms as a route to new antimicrobial, antitumor and neuroregenerative drugs. *Curr Opin Investig Drugs* **2** (12), 1681 (2001); Nguyen, K. T. et al., Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc Natl Acad Sci U S A* **103** (46), 17462 (2006).

21 Fischbach, M. A. and Walsh, C. T., Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106** (8), 3468 (2006).

22 Hubbard, B. K., Thomas, M. G., and Walsh, C. T., Biosynthesis of L-p-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. *Chem Biol* **7** (12), 931 (2000).

23 Weist, S. et al., Mutasynthesis of glycopeptide antibiotics: variations of vancomycin's AB-ring amino acid 3,5-dihydroxyphenylglycine. *J Am Chem Soc* **126** (19), 5942 (2004).

24 Stephen J. Hammond, Michael P. Williamson, Dudley H. Williams, LaVerne D. Boeck, and Gary G. Marconi, On the Biosynthesis of the Antibiotic Vancomycin. *J. Chem. Soc., Chem. Commun.*, 344 (1982).

25 Sandercock, A. M., et al., Biosynthesis of the di-meta-hydroxyphenylglycine constituent of the vancomycin-group antibiotic chloroeremomycin. *Chem. Commun.*, 1252 (2001).

26 Pfeifer, V. et al., A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine. *J Biol Chem* **276** (42), 38370 (2001).

27 Funa, N. et al., A new pathway for polyketide synthesis in microorganisms. *Nature* **400** (6747), 897 (1999).

28 Chen, H., Tseng, C. C., Hubbard, B. K., and Walsh, C. T., Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc Natl Acad Sci U S A* **98** (26), 14901 (2001).

29 Tseng, C. C., Vaillancourt, F. H., Bruner, S. D., and Walsh, C. T., DpgC is a metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase in the vancomycin biosynthetic Pathway. *Chem Biol* **11** (9), 1195 (2004).

30 Massey, V., Activation of molecular oxygen by flavins and flavoproteins. *J Biol Chem* **269** (36), 22459 (1994).

31 Joosten, V. and van Berkel, W. J., Flavoenzymes. *Curr Opin Chem Biol* **11** (2), 195 (2007).

32 Kelly, D.R., Wanm P.W.H., and Tang, J., *Biotechnology*. (Wiley, New York, 1998).

33 Sheng, D., Ballou, D. P., and Massey, V., Mechanistic studies of cyclohexanone monooxygenase: chemical properties of intermediates involved in catalysis. *Biochemistry* **40** (37), 11156 (2001).

34 Malito, E., Alfieri, A., Fraaije, M. W., and Mattevi, A., Crystal structure of a Baeyer-Villiger monooxygenase. *Proc Natl Acad Sci U S A* **101** (36), 13157 (2004).

35 Hecht, H. J. et al., Crystal structure of glucose oxidase from Aspergillus niger refined at 2.3 A resolution. *J Mol Biol* **229** (1), 153 (1993).

36 Fetzner, S., Oxygenases without requirement for cofactors or metal ions. *Appl Microbiol Biotechnol* **60** (3), 243 (2002).

37 Cane, D. E. and Walsh, C. T., The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chem Biol* **6** (12), R319 (1999); Cane, D. E., Walsh, C. T., and Khosla, C., Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282** (5386), 63 (1998).

38 Shen, B. and Hutchinson, C. R., Tetracenomycin F1 monooxygenase: oxidation of a naphthacenone to a naphthacenequinone in the biosynthesis of tetracenomycin C in Streptomyces glaucescens. *Biochemistry* **32** (26), 6656 (1993).

39 Sciara, G. et al., The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis. *Embo J* **22** (2), 205 (2003).

40 Beinker, P. et al., Crystal structures of SnoaL2 and AclR: two putative hydroxylases in the biosynthesis of aromatic polyketide antibiotics. *J Mol Biol* **359** (3), 728 (2006).

41 Bauer, I., Max, N., Fetzner, S., and Lingens, F., 2,4-dioxygenases catalyzing N-heterocyclic-ring cleavage and formation of carbon monoxide. Purification and some properties of 1H-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase from Arthrobacter sp. Ru61a and comparison with 1H-3-hydroxy-4-oxoquinoline 2,4-dioxygenase from Pseudomonas putida 33/1. *Eur J Biochem* **240** (3), 576 (1996).

42 Frerichs-Deeken, U. et al., Dioxygenases without requirement for cofactors and their chemical model reaction: compulsory order ternary complex mechanism of 1H-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase involving general base catalysis by histidine 251 and single-electron oxidation of the substrate dianion. *Biochemistry* **43** (45), 14485 (2004).

Chapter 2: Solving the apo-structure of DpgC: A cofactor-independent dioxygenase in the vancomycin biosynthetic pathway.

Introduction

It was established early in the study of DpgC that, based on the primary structure, the enzyme is homologous to the crotonase superfamily.[1] An enzyme superfamily is defined by a common structural strategy used to lower the free energies of chemically similar intermediates.[2] The crotonase superfamily is characterized by stabilization of coenzyme A acyl enolate intermediates (Figure 1). From these common intermediates, members of the superfamily catalyze a wide range of reactions, thought to emerge through divergent evolution.[3] In the case of enzymes, divergence is achieved by maintaining the functional groups that form the common intermediate, while adding and subtracting groups leading to various products.[4] The structural motif common to crotonase domains is an "oxyanion hole" formed by two backbone amide NH groups, which stabilizes the enolate anion form of acyl-CoA substrates.

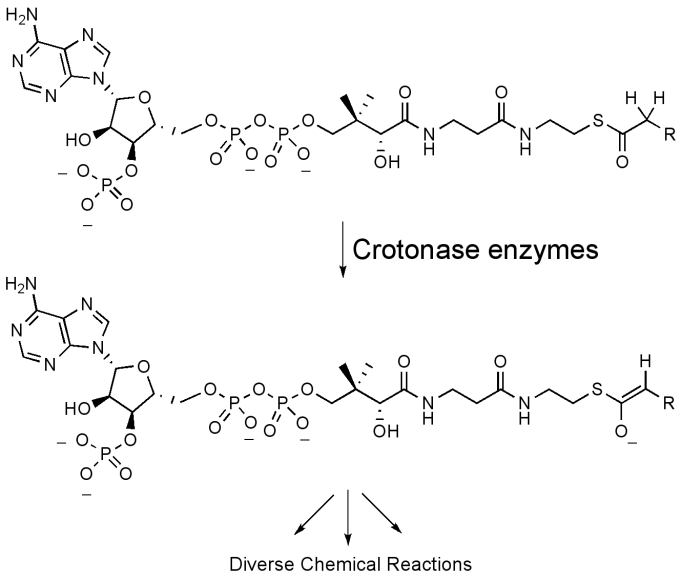


Figure 1: Initial catalytic step of the crotonase superfamily of enzymes

Crotonase enzymes catalyze a diverse array of reactions, including dehalogenation, hydration and dehydration, decarboxylation, formation and cleavage of carbon-carbon bonds and hydrolysis of thioesters (Figure 2). Several structures of

crotonase enzymes have been solved, which gives structural basis for the chemistry of this superfamily.[2]

Enoyl-CoA hydratase, referred to as crotonase from which the superfamily gets its name, catalyzes addition of water to α,β-unsaturated enoyl-CoA thioesters.[5] Crotonase is involved in the β-oxidation cycle of long chain fatty-acids. The Wierenga lab solved the crystal structure of crotonase bound to a substrate analog providing insight into how the enzyme performs its chemistry.[6,7]

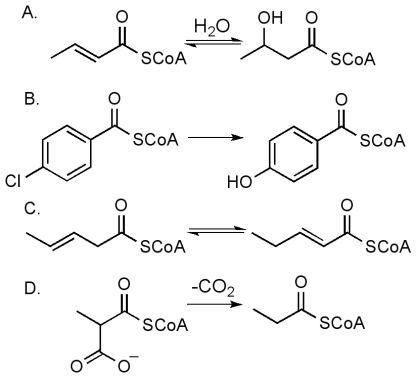


Figure 2: Reactions catalyzed by crotonase superfamily: A. hydratase B. dehalogenase C. isomerase D.decarboxylase

Glu144 of crotonase activates a water molecule that adds into the β-carbon of the substrate (Figure 3A). This use of a CoA enolate intermediate in the example of crotonase is more typical of the crotonase superfamily than 4-chlorobenzoyl CoA dehalogenase (Figure 3B).

4-chlorobenzoyl-CoA dehalogenase catalyzes the dechlorination and subsequent hydration of 4-chlorobenzoyl-CoA.[8,9] The Holden group solved the structure of this member of the crotonase superfamily bound to its product in order to elucidate the role of the oxyanion hole and enolate intermediate in this transformation.[10] The backbone amides of Phe64 and Gly114 form the oxyanion hole that holds the carbonyl of the thioester. Phe64, Trp89 and Trp137 form an aromatic and hydrophobic pocket around the 4-chlorobenzoyl moiety of the substrate. The oxyanion hole activates the C4 position for formation of the Meisenheimer complex with Asp145 (Figure 3B). This use of the oxyanion hole is quite different from hydratase enzymes of the crotonase superfamily.

DpgC represents the first known example of oxidation/reduction chemistry from the crotonase superfamily.[11] Similar to ActVA-Orf6 and SnoaL2 discussed in the first chapter of this thesis, DpgC demonstrates new reactivity from a diverse protein scaffold.[12,13] A crystal structure of DpgC would shed light on how the stabilization of an enolate intermediate activates the substrate for a four-electron oxidation and thioester cleavage. A crystal structure could also elucidate oxygen binding and activation.

Unfortunately the structure of DpgC unbound to an inhibitor or substrate mimic did not provide a complete picture of the active site chemistry. There were several disordered regions in the crystal structure where the electron density could not be interpreted. The apo-structure did provide some information concerning the overall fold of the enzyme and the interface between the monomers. The N-terminal ~third of the enzyme, which showed no sequence homology to any deposited protein sequences, was elucidated in this structure and was determined to be a novel fold.

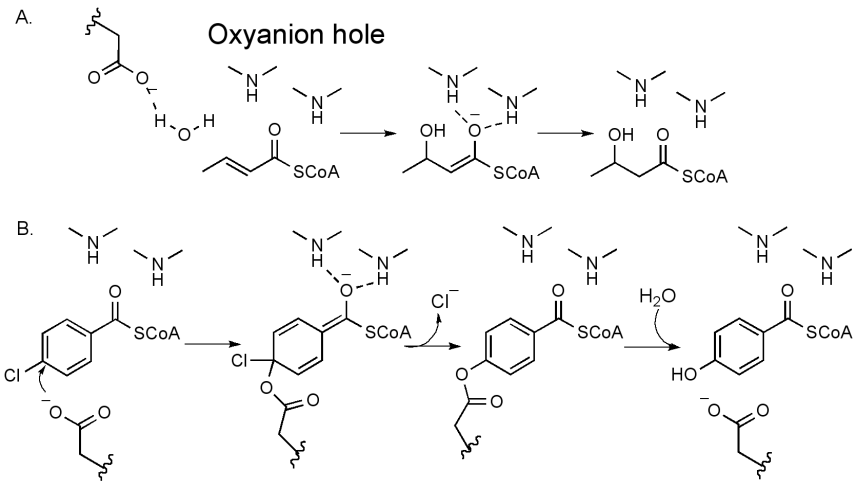


Figure 3: Proposed mechanisms for hydratase and dehalogenase

Overexpression and purfication of DpgC

DpgC from the A47934 gene cluster was cloned from cosmid DNA (provided by G.D. Wright, McMaster University) into the pET30a vector. Upon transformation into *Escherichia coli* and induction of overexpression, a significant amount of DpgC was produced, 10 mg/L of growth media. DpgC was cloned with a hexa-histidine tag attached to the N-terminus of the protein. This allowed for purification with Ni-NTA affinity resin. After elution from the resin, the hexa-histidine tag was cleaved from DpgC with the protease enterokinase. Successful crystallization requires protein to be purified to high homogeneity. To obtain DpgC suitable for crystallography, the enzyme was furthered purified using a two-step protocol: anion-exchange column followed by size exclusion chromatography. This provided DpgC in high enough purity for high-throughput crystallography screens.

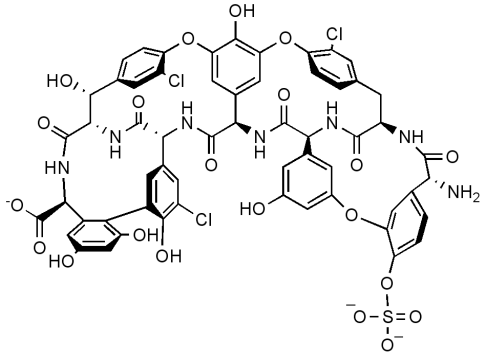


Figure 4: The vancomycin family member A47934

Crystallization and initial X-Ray data collection of DpgC

Work on crystallizing DpgC began with screening several DpgC homologues from various vancomycin family biosynthetic pathways against high-throughput crystal screens. Due to the high sequence homology, approximately 70% identity and 80% similarity, crystallization of any DpgC homolog should provide a model applicable to the entire family of enzymes. After the purification procedure described above, DpgC from A47934 was found to crystallize in 95 mM NaHEPES, 0.94 M NaCl and 1.3 M

$(NH_4)_2SO_4$, pH 7.5. Data was collected on these crystals at our home source, a rotating copper anode X-ray generator. Analysis of diffraction data showed DpgC crystallizes in the *R3* space group with unit cell dimensions of a,b=138.7 Å, c=239.7 Å, $\alpha=\beta=90^o$, $\gamma=120^o$. Based on the Matthews coefficient, we predicted each asymmetric unit would contain four DpgC monomers.

Determining phase information

X-ray diffraction data of a macromolecule cannot be used to generate an electron density map in the absence of calculated phase information. The reflections seen in an X-ray diffraction pattern are the result of a wave of X-ray radiation diffracting by interaction with clouds of electrons in a protein crystal then measured with a detector. Each reflection, or spot, on a diffraction pattern is made up of contributions from all scattering elements in a crystal.[14] Like any wave, X-rays diffracted by a protein crystal can be described by three parameters: wavelength, amplitude and phase. Amplitude of the X-ray is the square root of the reflection intensity measured from the diffraction pattern. The wavelength of the X-rays diffracted by a crystal is constant with the diffraction source. Phase information, however, cannot be obtained through direct analysis of diffraction data; this must be calculated through indirect methods. With data describing amplitude, wavelength and phase, the wave function that represents the electron density can be solved through a Fourier transform. To obtain the missing piece of data and solve the "phase problem," three experimental methods are commonly used.

When a highly homologous model of the protein of interest is available a method called molecular replacement can be employed. In this case the homologous model is

used as a search model for the unknown phase information. The phases from the structure factors of the search model are used to make initial estimations for the phases of the new protein.

A second method involves soaking the protein of interest with a heavy atom salt then crystallizing the protein, or soaking an already formed crystal of the protein of interest with a heavy atom salt. In this procedure, several heavy atom salts are screened in hopes of finding one that will bind consistently to the protein in several locations. The diffraction data from this heavy atom derivative of the protein has structure factors from the heavy atom (F_H) as well as the structure factors from the protein (F_P; $F_{HP} = F_H + F_P$). Subtracting the data of a native crystal from the data of a heavy atom derivative crystal leaves structure factors only from the bound heavy atoms ($F_H = F_{HP} - F_P$). Computational methods can be used to determine the location of the bound heavy atoms, creating phase information that can then be applied to the rest of the protein. This approach has the disadvantage of requiring data sets from separate crystals. Because of the imperfections in a macromolecular crystal, data from two crystals of the same protein are not strictly identical, creating a larger error factor when comparing the two data sets.

The third method of phase determination is anomalous dispersion, which relies on an atom in the protein having an anomalous diffraction edge. In this case data collection is performed at certain wavelengths that create an anomalous signal upon X-ray diffraction by the atom. Anomalous data is generated because certain atoms absorb X-rays of certain wavelengths anomalously causing inequality in symmetry related reflections. Data is collected as either single-wavelength anomalous dispersion (SAD), or multi-wavelength anomalous dispersion (MAD). In the case of MAD, multiple data

sets are collected on a single crystal at different wavelengths. These sets are then compared to generate phase information. SAD collects a single data set at the absorption edge of the heavy atom incorporated into the protein. This provides adequate data to determine phase information. SAD has recently become the common method for anomalous data collection. Powerful software has made the collection of multiple data sets increasingly unnecessary.[15] Analysis of the anomalous data allows for the atoms to be located, and phase information to be calculated. Selenium atoms have an absorption edge within the range of synchrotron radiation. This was the first heavy atom incorporated into DpgC.

DpgC heavy atom soaks: MIR and anomalous data collection

DpgC crystals were soaked with a range of heavy atom salts, including $HgCl_2$, HgOAc, NH_4PtCl_4, $KPtCl_6$. Several datasets were collected on the heavy atom soaked crystals on our home source. Patterson maps of the derivative crystals were made in an attempt to derive phase information. Peaks observed in Patterson maps represent vectors between atoms. The observed Patterson map shown is based on a subtraction of a native data set from a $HgCl_2$ derivative DpgC crystal (Figure 5). This subtraction should result in peaks representing vectors only between heavy atoms in the crystal. A calculated Patterson map is shown along with an overlay of the two maps. The possible correlation between the calculated and observed map gave some hope that the crystal had bound Hg atoms in an ordered and consistent fashion. Attempts to use this data to determine phase information through the MIR strategy and calculate an electron density map were unsuccessful. In order to determine whether DpgC had bound any heavy atoms, crystals

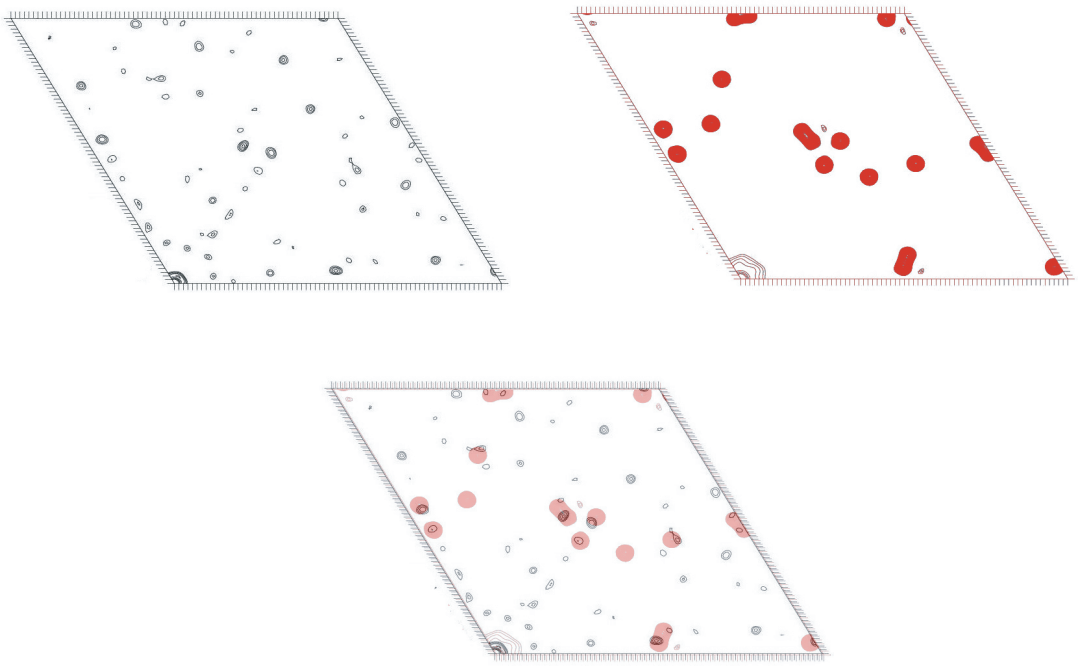


Figure 5: Upper left, experimentally determined Patterson map: F_{HP}-F_P=H_H. Upper right, calculated Patterson map. Overlay of experimental and calculated, lower middle

were sent to the National Synchrotron Light Source (NSLS) to collect anomalous data. Scans of the crystals showed very weak anomalous signal (Figure 8). Data sets were collected at the absorption edge for the specific heavy atom. None of the derivative crystals gave a strong enough anomalous signal to provide phase information. At this point it became necessary to pursue an alternate strategy towards phase determination.

Initial SAD phasing experiments of DpgC

Once crystallization conditions for DpgC were optimized for diffraction quality, the enzyme was prepared in minimal media incorporating seleno-methionine (SeMet) into the peptide chain. The strategy of a SeMet protein prep is to deprive the producing organism of methionine while supplying an excess of SeMet. The media for a SeMet preparation is composed of necessary salts, including ammonium salts for nitrogen source, glucose as a carbon source and necessary amino acids. Cysteine and other amino acids in the biosynthetic pathway towards methionine are excluded to keep methionine production as low as possible. This SeMet prep produced a comparable amount of DpgC to the native prep. Crystallization of the SeMet-substituted enzyme was successful using the native conditions. However, the SeMet derivative enzyme crystallized in a different space group from the native, *P32*, and the data could not be merged. Crystal statistics produced from data reduction indicated possible twinning. Crystal twinning is a phenomenon in which the entire asymmetric unit of a crystal adopts multiple conformations within the crystal lattice. This results in each reflection in the diffraction pattern coming from two separate asymmetric units. In some cases this data can be deconvoluted, however we were unsuccessful in any attempt to interpret the SeMet-DpgC data. A possible explanation for the crystal twinning is the high number of methionines, ten, in DpgC. Because selenium is easily oxidized, it is possible that a significant number of SeMet in DpgC became oxidized (Figure 6) and disrupted the formation of a properly ordered crystal. Once again, another strategy for phase determination was employed.

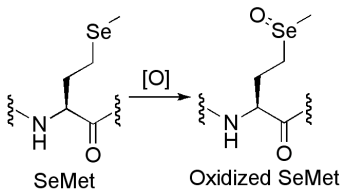


Figure 6: Oxidized SeMet

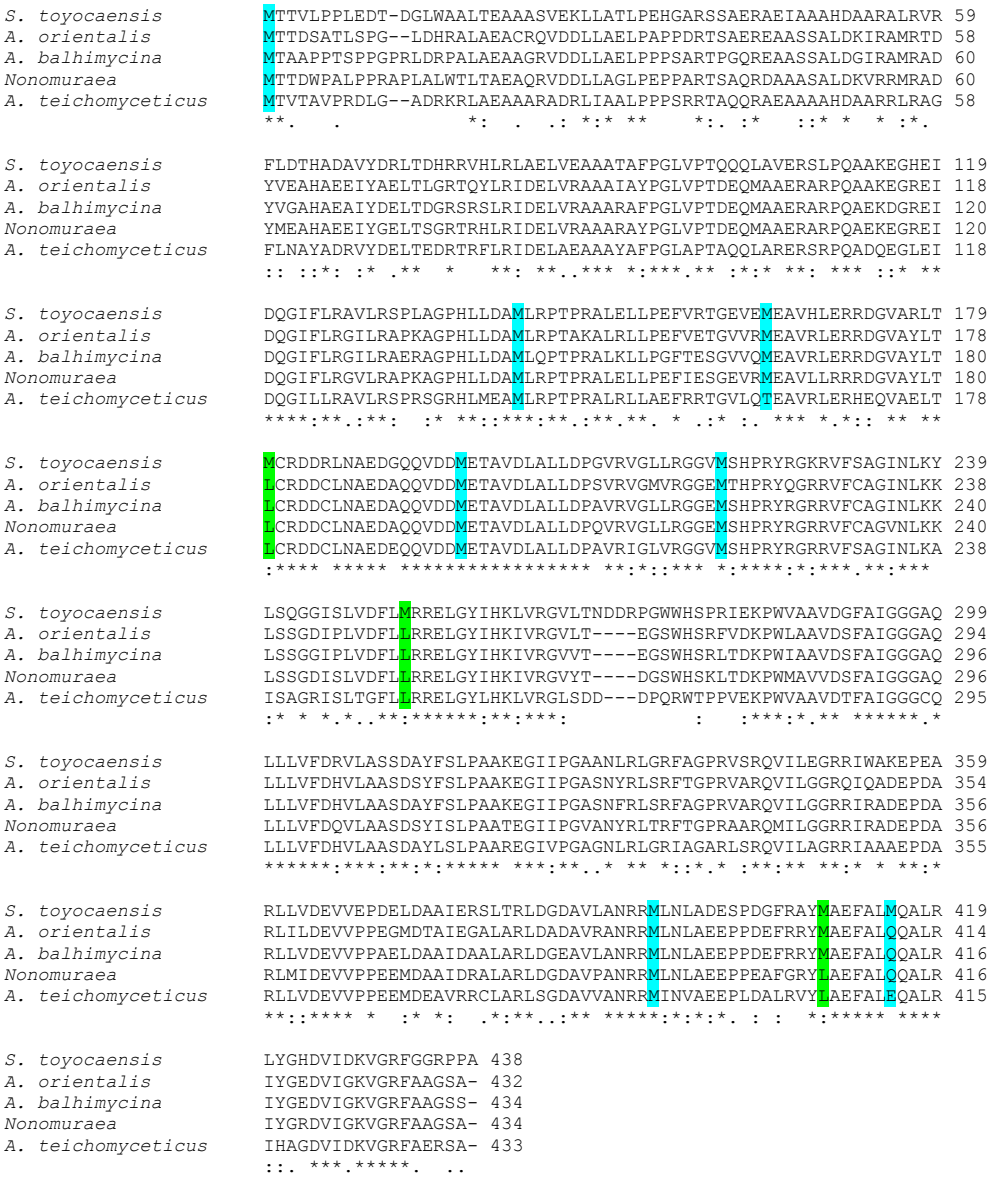


Figure 7: Alignment of DpgC from five different glycopeptides antibiotic clusters. DpgC from the A47934 pathway is listed first. Teal highlights methionines not mutated in the triple mutant. Green highlights methionines mutated to leucine

SAD data collection on a triple mutant of DpgC, and phase determination

We believed DpgC produced twinned SeMet-derivative crystals due to the high (ten) number of methionines in the enzyme; therefore three methionines were mutated to leucine. This strategy seems counterintuitive as removing methionines from a protein means reducing the anomalous signal from a SeMet derivative, therefore reducing the phasing power of the crystallographic data. However, if such a mutant could lead to crystals that were not twinned, the remaining methionines should provide adequate phasing power to solve the structure.

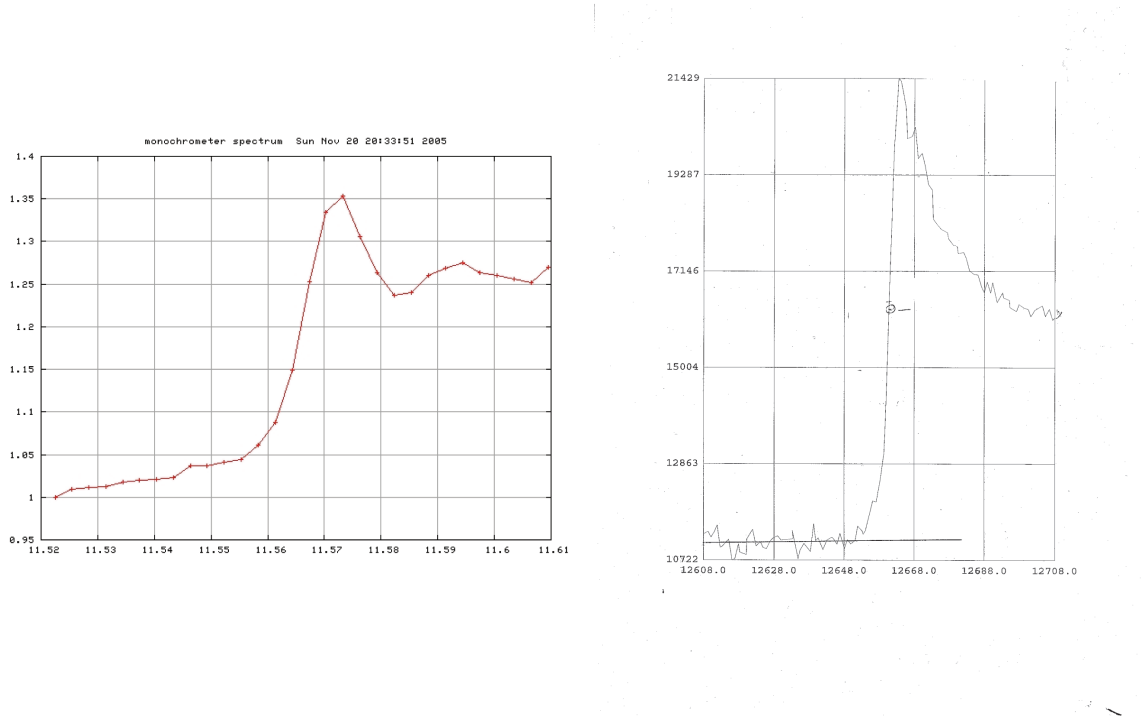


Figure 8: Anomalous energy scans. SeMet DpgC triple mutant on right, $HgCl_2$ derivative DpgC on left Fluorescence radiation (relative scale on Y-axis) is measured as X-ray wavelength (in eV on x-axis) is adjusted. The sharp increase in from the SeMet DpgC radiation indicates an anomalous signal powerful enough for phasing.

A sequence alignment of DpgC homologues showed three of the methionines in A47934 DpgC were not conserved (Figure 7). In all three of those positions, leucines were present in at least two DpgC homologues. The three green-highlighted methionine

residues were mutated to leucine using the Stratagene QuikChange mutagenesis kit, M180L, M252L and M409L (Figure 7). The triple mutant crystallized in the *R3* space group and was not twinned according to analysis of data from our home source X-ray generator. Anomalous data were then collected on the X4A beamline at NSLS. Phasing calculations were carried out using the PHENIX[16] software package, and 19 of a possible 28 selenium atoms were located, with an overall figure of merit of 0.26. Initial electron density maps were calculated using the SAD phases from PHENIX to 2.75 Å. With electron density maps derived from the SAD phases, it was possible to begin building a model of DpgC.

Building the structure of apo-DpgC

The model of DpgC was built using the program COOT[17] with structure refinement by the CNS software package.[18] DpgC contains four monomers in the asymmetric unit. Each monomer forms a trimer through symmetry operations. One of the four monomers displayed very poor electron density and could not be built beyond a largely incomplete polyserine model. Two monomers had notable higher quality density and gave the most complete models.

Unfortunately, even this monomer could not be completed due to several disordered regions (Figure 10). Refinement statistics for the apo-DpgC model did not reach a level worthy of publication. Although some structural knowledge of the enzyme was gained from this model, little could be deciphered about the chemistry of the enzyme.

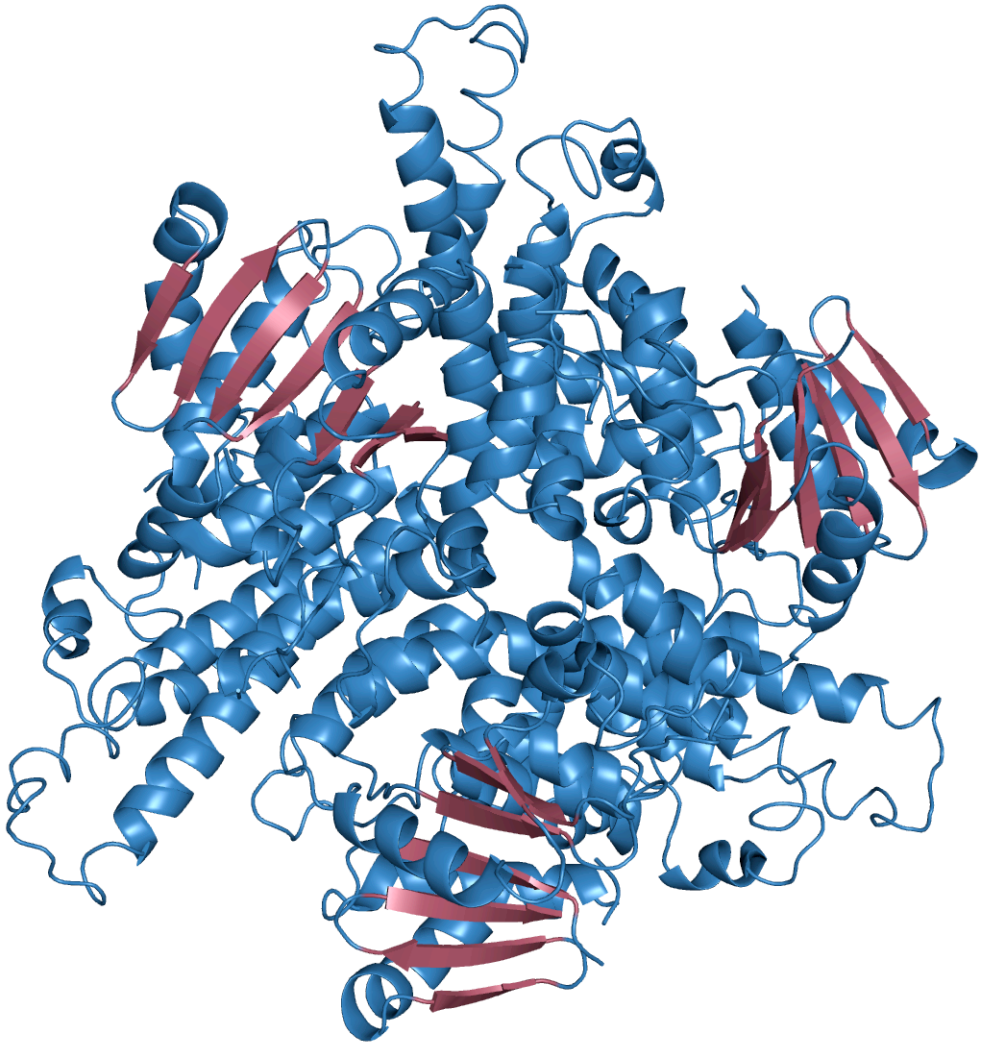
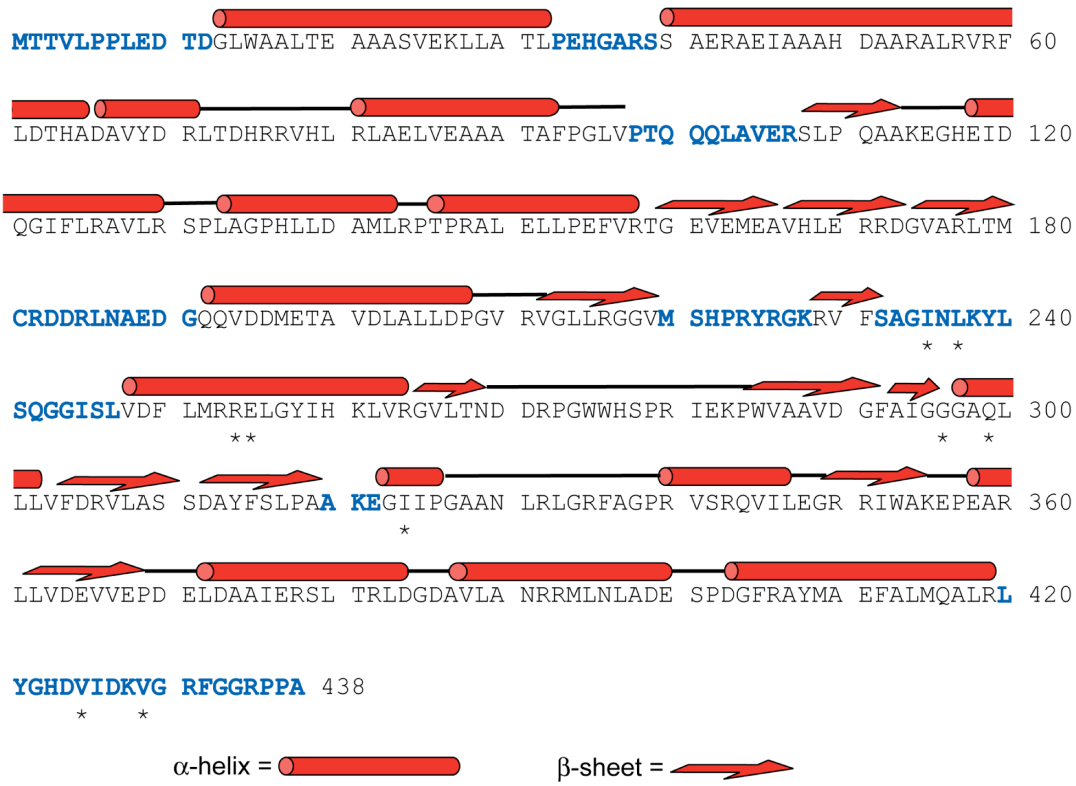


Figure 9: DpgC looking down a three-fold axis of symmetry. The biological unit of crotonase domains is a dimer of trimers

The biological unit of most crotonase enzymes is a dimer of trimers (Figure 9).[2] DpgC crystallized as four monomers, each monomer forms a trimer through symmetry operations. Therefore, the biological unit of DpgC could be observed in this initial structure. Hydrophobic α-helices make contacts between neighboring monomers. A fascinating aspect of the structure is the structural characterization of the N-terminal ~third of the enzyme. The amino acid sequence of this portion is not homologous to any

A.



B.

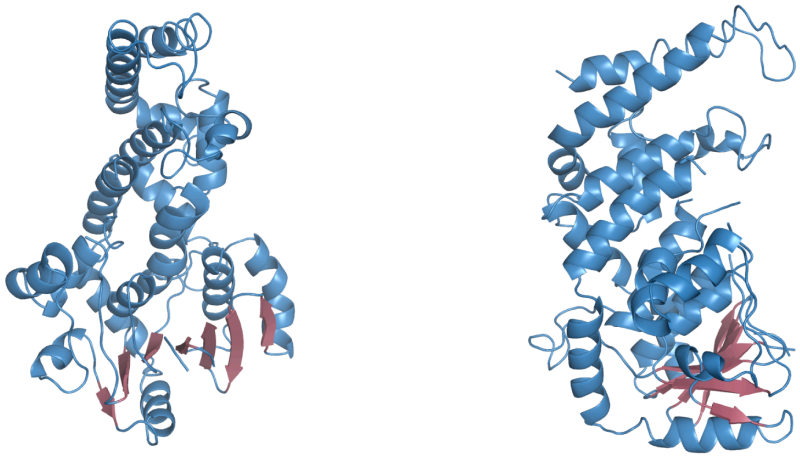


Figure 10: A. Amino acid sequence of DpgC with cartoon representations of α–helices and β–sheets above the sequence. Highlighted in blue are residues that could not be built. Asterisks indicate catalytically relevant residues. B. Two perspectives of DpgC monomer.

other known sequence, likewise the structure of this sequence shows no homology when put through a DALI structure homology search. The C-terminal two-thirds of the

structure is largely homologous to crotonase domains in terms of its tertiary and quaternary structure. In spite of this, elucidating the chemistry of the enzyme remains a mystery due to the poor electron density surrounding the active site.

Crystallography depends on a protein existing in the same conformation in every asymmetric unit in the crystal. The electron density in an area of disorder is either non-existant or random. As opposed to twinned data, which is the result of entire asymmetric

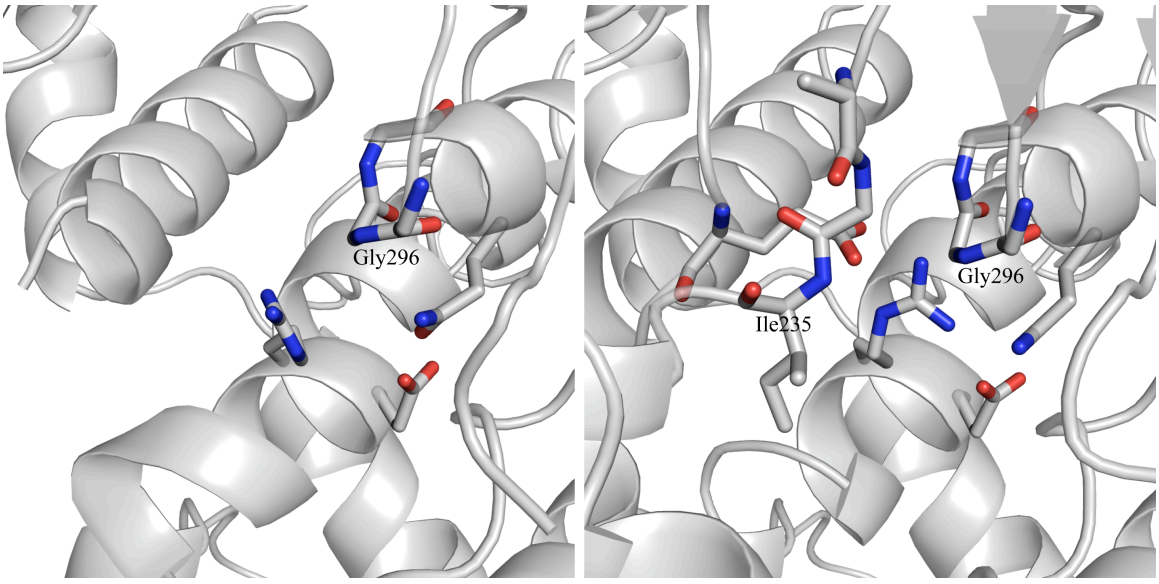


Figure 11: Apo-DpgC showing an incomplete active site on the left. Gly296 represents one of the two residues composing the oxyanion hole. The complete structure from DpgC bound to substrate mimic, on the right. The complete oxyanion hole is built, Ile235 and Gly296. Not shown: the substrate mimic

units in different conformations, the data of apo-DpgC had all symmetric units properly aligned. The areas surrounding the DpgC active site, as well as other portions on the enzyme, adopt multiple conformations in the crystal. If a protein is in multiple conformations within a crystal, X-ray diffraction will not lead to a clear image of the protein. Previous studies of oxygenase enzymes indicate the substrate binds the active site followed by molecular oxygen.[19,20] If a substrate mimic or inhibitor were bound to

the protein, a more stable complex might form. This stable complex might provide a structural basis for substrate and molecular oxygen binding to the protein.

DpgC unbound to a substrate mimic does not even provide a picture of the oxyanion hole in the active site (Figure 11). When compared to the crystal structure bound to the substrate mimic, which will be discussed in the next chapter, it is clear that only one of the amides in the oxyanion hole is ordered. This leaves us with little information on one of the fundamental aspects of crotonase enzymes, as well as a crucial aspect of the chemistry of DpgC.

Materials and Methods

DpgC cloning and overexpression

DpgC from the A47934 gene cluster was cloned from cosmid DNA (provided by G.D. Wright, McMaster University) into the NcoI and HindIII sites of the pET30A vector. The vector was transformed into BL21(DE3) cells and grown in LB media at 37^{o}C until cell density reached O.D.=0.5-0.7. Overexpression was induced by adding 100 μL 0.5 M IPTG, followed by overnight incubation at 18^{o}C. Cells were pelleted by centrifugation at 3,500 rpm for 20 minutes and lysed by passage through a French Press cell disruption system at 1,000 psi. The cell lysate was incubated with Ni-NTA affinity resin (Qiagen) for 1 hr. The resin was washed four times with 10 mL 500 mM NaCl, 20 mM TrisHCl, pH 7.5. The enzyme was eluted from the resin by washing two times with 10 mL 500 mM NaCl, 250 mM imidazole, 20 mM TrisHCl, pH 7.5. The elutions were dialyzed against 1 L of 100 mM NaCl, 20 mM TrisHCl, 1 mM βME, pH 7.5, followed by cleavage of the hexa-histidine tag with the protease enterokinase (2 days, 4^{o}C). The enzyme was further purified with a HiTrap-Q (GE Biosciences) ion exchange column (buffer A: 20 mM Tris, 1 mM βME, pH 7.5, buffer B: 1 M NaCl, 20 mM TrisHCl, 1 mM βME, pH 7.5) flow rate 2 mL/min, gradient 0-80% buffer B over 60 mL. The enzyme was purified to homogeneity with a HiLoad 16/60 SuperDex 200 gel filtration column (GE Biosciences), buffer 100 mM NaCl, 20 mM TrisHCl, 1 mM βME, pH 7.5. Flow rate 1mL/min over 150 min, DpgC elutes around 60 min.

Crystallization of DpgC

DpgC was crystallized by the hanging drop method at 20 ^{o}C. DpgC (1.5 μL of 10 mg/mL DpgC in 20 mM Tris, 100 mM NaCl, pH 7.5) was mixed with 1.5 μL reservoir solution (95 mM HEPES, 0.94 M NaCl, 1.3 M $(NH_4)_2SO_4$, pH 7.5). Crystals appeared after two days. Crystals were transferred to a cryoprotectant solution, composed of crystallization solution supplemented with 20% glycerol, and soaked for 10 min. before being flash frozen in liquid nitrogen.

DpgC seleno-methionine prep

To M9 media (6 g Na_2HPO_4, 3 g KH_2PO_4, 1 g NH_4Cl, 0.5 g NaCl in 1 L dd H_2O) was added, 30 mg kanamycin, 10 mL 20% glucose, 2 mL 1M $MgSO_4$, 0.05 mL 2M $CaCl_2$, 0.1M 0.5% weight/volume (w/v) thiamine; 40 mg of the following amino acids: arginine, phenylalanine, tyrosine, tryptophan, lysine, histidine, serine, glutamate, aspartate, threonine, leucine, isoleucine and valine. This mixture is inoculated with 1 mL of fresh overnight DpgC culture in LB broth, and grown to O.D.=0.5-0.7. At this point 100 mg each of threonine, lysine, phenylalanine and 50mg leucine, isoleucine and valine, and 120 mg DL-SeMet were added to media. The culture is shaken for 15 min during which it is cooled to 18^{o}C. 100 μL of 0.5 M IPTG is added, and the culture is shaken for 16 hr at 18 ^{o}C.

Purification of the SeMet prep follows the protocol described above with a few exceptions. All buffers were degassed for 10min with argon. All buffers were identical except for containing 10 mM DTT. SeMet-DpgC crystallized under the same conditions as native DpgC.

DpgC heavy atom soaks

DpgC crystals were soaked in 95 mM HEPES, 0.94 M NaCl, 1.3 M Li_2SO_4, pH 7.5, several times to remove any $(NH_4)_2SO_4$ as the ammonium ion would bind to the metal salts competing with DpgC. Heavy atoms salts $HgCl_2$, $HgOAc_2$, $KPtCl_6$, NH_4PtCl_4, $KPtCN_4$, $PbNO_3$ and $KAu(CN)_2$ were screened at concentrations of 1-5mM in the above mentioned solution. Crystals were soaked in the heavy atom solution from 0.5 hr. to 16 hr. Data on the heavy atom soaked crystals was collected on our home source as well as on X26C at the NSLS. No data sets yielded phasing power or anomalous signal.

Site-directed mutagenesis of DpgC

Point mutations of DpgC were constructed using the QuikChange Multi Site-Directed Mutagenesis Kit (Stratgene). PCR gene amplification was performed using components provided; QuikChange Reaction Buffer, QuikChange Multi enzyme blend and dNTP mix. The following primers and their reverse compliments were used for the triple mutant (modified sequences underlined): **M180L**, 5'-CGG CTG ACC <u>CTG</u> TGT CGC GAC GAC CGC-3'; **M252L**, 5'-GAC TTC CTG <u>CTG</u> CGC CGG GAA CTC GGC-3'; **M409L**, 5'-CGC GCG TAC <u>CTG</u> GCG GAG TTC GCC CTC-3'. The PCR program consisted of an initial hold of 95^{o}C for 1 min followed by 30 cycles of 95^{o}C for 1min, 55^{o}C for 1 min, and 65^{o}C for 13.5 min. The template DNA was digested with 10U of DpnI for 1 hour at 37^{o}C before transformation into XL1-Gold Ultracompetent Cells. The mutagenesis products were confirmed by DNA sequencing (Genewiz, Inc.).

DpgC structure determination using single-wavelength anomalous dispersion

The SeMet-DpgC triple mutant crystallized under native DpgC crystallization conditions. Anomalous data was collected on the X4A beamline at NSLS, with a wavelength at the Se edge (0.98088 Å). Diffraction data were indexed, integrated and scaled using the HKL2000 program. Heavy atom phasing was carried out with the PHENIX[16] software package.

Model construction based on the SAD derived electron density maps

The model was built with the program COOT[17] and refined with the software suite CNS.[18] Initial structural refinement was performed with non-crystallographic symmetry (NCS) restraints, after several rounds the restraints were removed from calculations. Sigma-weighted simulated annealing composite omit maps (CNS) were used to build the model. Issues completing the model are discussed in the text of chapter 2.

Table 1: Crystallography data collection and refinement statistics

Data collection	SeMet DpgC	apo-DpgC
Space group	R3	R3
Cell dimensions		
a, b, c (Å)	138.7, 138.7, 239.7	138.7, 138.7, 240.7
α, β, γ (°)	90, 90, 120	90, 90, 120
Wavelength	0.98088	1.000001
Resolution (Å)	2.75	2.21
R_{sym} or R_{merge}	0.061 (0.714)	0.085 (0.821)
$I/\sigma I$	13.94 (2.47)	26.37 (2.00)
Completeness (%)	98.4 (99.1)	99.1 (95.5)
Redundancy	3.6 (3.6)	3.0 (2.8)
Refinement		
Resolution (Å)	2.75	2.21
No. reflections	88780	86521
R_{work}/R_{free}		0.319/0.330
No. atoms		11099
Average B-factor		59.5
R.M.S.D.		
Bond lengths (Å)		0.008
Bond angles (°)		1.5

References:

1 Pfeifer, V. et al., A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine. *J Biol Chem* **276** (42), 38370 (2001).

2 Holden, H. M., Benning, M. M., Haller, T., and Gerlt, J. A., The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme a thioesters. *Acc Chem Res* **34** (2), 145 (2001).

3 Gerlt, J. A. and Babbitt, P. C., Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr Opin Chem Biol* **2** (5), 607 (1998).

4 Babbitt, P. C. and Gerlt, J. A., Understanding enzyme superfamilies. Chemistry As the fundamental determinant in the evolution of new catalytic activities. *J Biol Chem* **272** (49), 30591 (1997).

5 Waterson, R. M. and Hill, R. L., Enoyl coenzyme A hydratase (crotonase). Catalytic properties of crotonase and its possible regulatory role in fatty acid oxidation. *J Biol Chem* **247** (16), 5258 (1972).

6 Engel, C. K. et al., Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 angstroms resolution: a spiral fold defines the CoA-binding pocket. *EMBO J* **15** (19), 5135 (1996).

7 Engel, C. K., Kiema, T. R., Hiltunen, J. K., and Wierenga, R. K., The crystal structure of enoyl-CoA hydratase complexed with octanoyl-CoA reveals the structural adaptations required for binding of a long chain fatty acid-CoA molecule. *J Mol Biol* **275** (5), 847 (1998).

8 Scholten, J. D. et al., Novel enzymic hydrolytic dehalogenation of a chlorinated aromatic. *Science* **253** (5016), 182 (1991).

9 Liang, P. H., Yang, G., and Dunaway-Mariano, D., Specificity of 4-chlorobenzoyl coenzyme A dehalogenase catalyzed dehalogenation of halogenated aromatics. *Biochemistry* **32** (45), 12245 (1993).

10 Benning, M. M. et al., Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 Å resolution: an enzyme catalyst generated via adaptive mutation. *Biochemistry* **35** (25), 8103 (1996).

11 Tseng, C. C., Vaillancourt, F. H., Bruner, S. D., and Walsh, C. T., DpgC is a metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase in the vancomycin biosynthetic Pathway. *Chem Biol* **11** (9), 1195 (2004).

12 Sciara, G. et al., The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis. *Embo J* **22** (2), 205 (2003).

13 Beinker, P. et al., Crystal structures of SnoaL2 and AclR: two putative hydroxylases in the biosynthesis of aromatic polyketide antibiotics. *J Mol Biol* **359** (3), 728 (2006).

14 Rhodes, Gale, *Crystallography Made Crystal Clear*, second ed. (Academic Press, San Diego, 2000).

15 Robinson, Howard, 2006.

16 Adams, P. D. et al., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58** (Pt 11), 1948 (2002).

17 Emsley, P. and Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60** (Pt 12 Pt 1), 2126 (2004).

18 Brunger, A. T. et al., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54** (Pt 5), 905 (1998).

19 Roth, J. P. and Klinman, J. P., Catalysis of electron transfer during activation of O2 by the flavoprotein glucose oxidase. *Proc Natl Acad Sci U S A* **100** (1), 62 (2003).

20 Frerichs-Deeken, U. et al., Dioxygenases without requirement for cofactors and their chemical model reaction: compulsory order ternary complex mechanism of 1H-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase involving general base catalysis by histidine 251 and single-electron oxidation of the substrate dianion. *Biochemistry* **43** (45), 14485 (2004).

Chapter 3: Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis: Crystal structure of DpgC bound to an isosteric substrate mimic.

Introduction

Aerobic organisms use molecular oxygen for critical processes such as energy production and synthesizing cellular components. The standard reduction potential of the O_2/H_2O redox couple is +0.82V, making dioxygen a potentially strong oxidizing agent.[1] However, as discussed in chapter 1, molecular oxygen cannot directly react with the vast majority of organic substrates due to the conservation of angular momentum. Nature has developed several methods to overcome this spin-forbidden reaction. The vast majority of solutions found to the problem of inert triplet oxygen come from organic cofactors or enzyme bound transition metals.[2]

Mechanistic studies of enzymatic molecular oxygen activation have focused on metalloenzymes. In general, this body of work has shown metalloenzymes both perform electron transfer and electrostatic stabilization in a single step.[3] Ligands bound to the metal center tune redox potentials of the systems towards increased reactivity with O_2.[4] The adjustable metal centers, single step electron transfer and charge stabilization are advantages that flavin-dependent enzymes do not have access to when activating oxygen. Instead flavoenzymes use specialized protein environments to facilitate charge transfer.[5] Flavin enzymology is frequently used as a guide to cofactor-independent oxygenation, as the substrate in cofactor-independent oxygenases is thought to function analogously to flavin cofactors.[1]

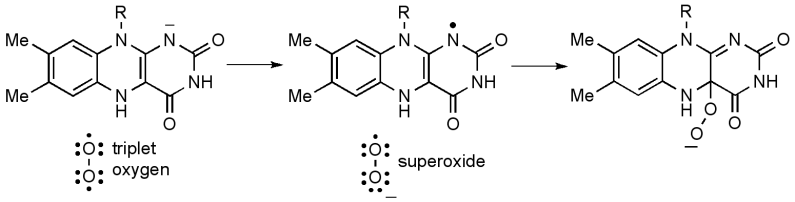


Figure 1: Single electron donation from reduced flavin to molecular oxygen

Catalytic rates of O_2 activation by flavin-dependent oxidases correlate to the presence of positively charged residues in the active site. Roth *et al.* have shown the importance of a positively charged residue in the active site of the flavin-dependent enzyme glucose oxidase (GO).[6] Through spectroscopic and kinetic analysis of this enzyme, the rate-limiting step was determined to be single electron transfer from the reduced flavin anion to molecular oxygen (Figure 1). Kinetic analysis of GO activity over a pH range shows rate acceleration at low pH, supporting the importance of a positive charge in the active site. In the case of GO, a positively charged histidine in the active site was shown to dramatically increase the rate of reaction. Roth *et al*. suggest a $HisH^+$ in the active site may stabilize $O_2^{\bullet-}$ (superoxide), which would increase the overall reaction driving force. The presence of charged amino acids are also implicated in creating the proper protein dielectric for electron transfer. The kinetic data by Roth *et al.* show a simple process of oxygen diffusion to the bound flavin, followed by single electron transfer. The importance of a charged residue in the active site led to the conclusion that enzymes using nonmetal cofactors for molecular oxygen activation are likely to use charged amino acids to facilitate electron transfer and stabilize intermediates.

Such studies of flavin-dependent enzymes provide clues on how the active site of DpgC might facilitate electron transfer and stabilize $O_2^{\bullet-}$. The DpgC active site has to provide an environment that will stabilize a negatively charged substrate and be suitable for electron transfer. The active site will also need a hydrophobic pocket to bind and orient molecular oxygen for reaction with the organic substrate.

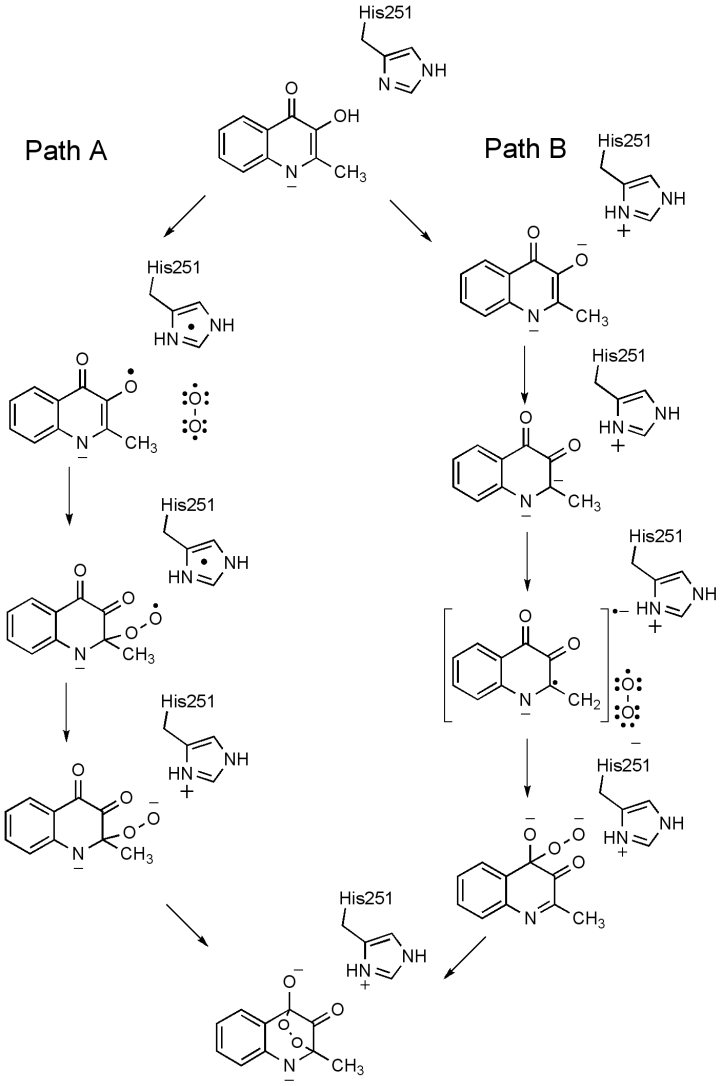


Figure 2: Possible mechanisms of cofactor independent dioxygenation by Hod. The use of a histidinyl radical is thought to be unlikely.

Two mechanistic routes concerning cofactor-independent dioxygenation have been suggested, in the case of Hod (1*H*-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase).[7] The first (Path A, Figure 2) involves hydrogen atom abstraction from the substrate to form a radical amino acid species, a histidinyl radical, in the active site. The substrate radical then reacts with molecular oxygen, as they have become spin-paired species. Next, single electron transfer occurs from the radical amino acid to the

hydroperoxyl radical species. Hydrogen atom abstraction is typically observed in metalloenzymes, however this has not been described in the case of organic cofactors. In fact, a histidinyl radical has been observed in bovine copper-zinc superoxide dimutase.[8] However, there are no known examples of catalytically competent histidinyl protein radicals.[9] The more likely path of cofactor independent dioxygenation is single electron transfer from a deprotonated substrate to molecular oxygen (Path B, Figure 2). The caged radical anion/superoxide pair would collapse to form a carbon-oxygen bond, followed by conversion to product, as shown. This mechanism is similar to the suggested flavin activation.

The work of the Fetzner and Klinmann labs provides insight towards how enzymes facilitate single electron transfer from an electron rich organic intermediate. We have addressed the issues of substrate activation, oxygen binding and active site environment leading to single electron transfer are addressed by solving the crystal structure of the cofactor-independent dioxygenase DpgC bound to an isosteric substrate mimic. DpgC is the first example of a crystal structure from this class of enzymes. Through this work we have developed a clear picture of the starting complex between DpgC and its substrate.

Rational design and synthesis of an isosteric substrate mimic for DpgC

DpgC catalyzes the four-electron oxidation of the α–C position of DpaCoA and cleaves the thioester bond to release coenzyme A and dihydroxyphenylglyoxylate (DpgX).[10] To probe the powerful chemistry of DpgC, we designed a stable isosteric substrate mimic. If cleavage of the thioester bond is prevented, the substrate will remain

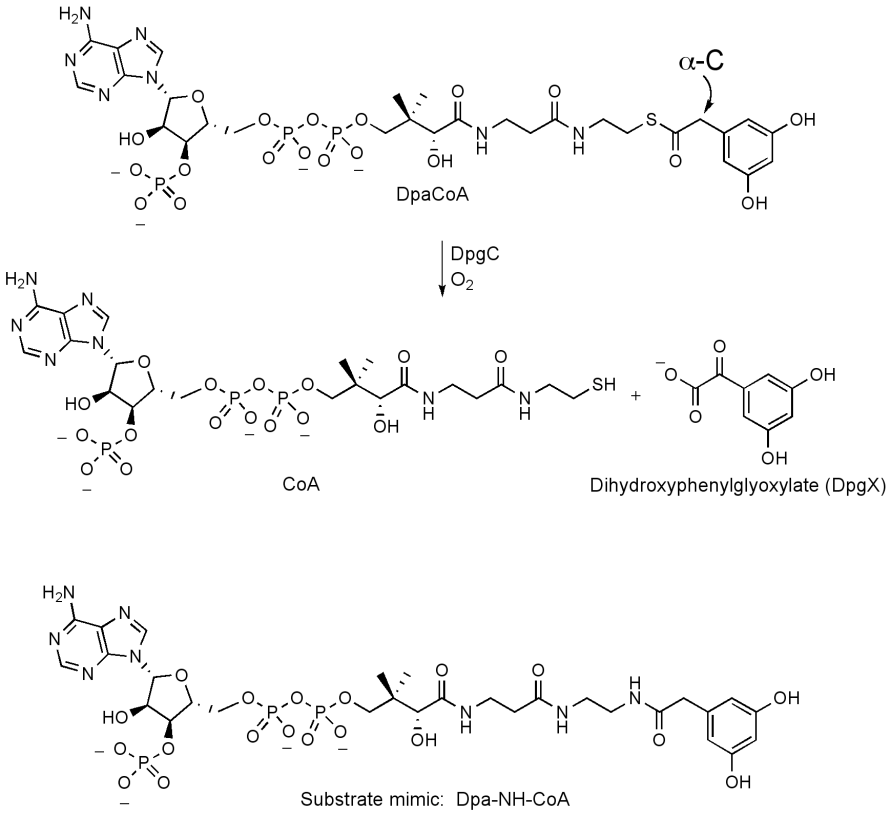


Figure 3: Transformation performed by DpgC, substrate mimic of DpgC

bound in the active site and provide a picture of the substrate-enzyme complex. It was not feasible to engineer the enzyme to prevent thioester cleavage; thus it became necessary to design a substrate analog that would be hydrolytically stable at this position. An amide bond is isosteric to a thioester with the exception of a single hydrogen. Compared to thioesters, amides are more resistant to hydrolysis. Another advantage of replacing the thioester of the substrate with an amide is preventing deprotonation at the α–C position, by raising the pKa. Because amide α-protons have a higher pKa than thioester α-protons, it was reasoned that the amide analog might prevent deprotonation at this location. Protonation at the C2 position would provide a true substrate complex with the enzyme, suggesting a substrate mimic with a change of simply one atom, sulfur to nitrogen. To synthesize such an inhibitor it is necessary to construct a coenzyme A analog with a terminal amine in place

of a thiol. The biosynthesis of coenzyme A played a large role in our synthetic strategy for the substrate mimic, leading to a chemoenzymatic approach.[11]

The pantetheine portion of coenzyme A is biosynthesized from a sequential coupling of pantoate, β-alanine and cysteine.[12] Pantetheine is then phosphorylated on the

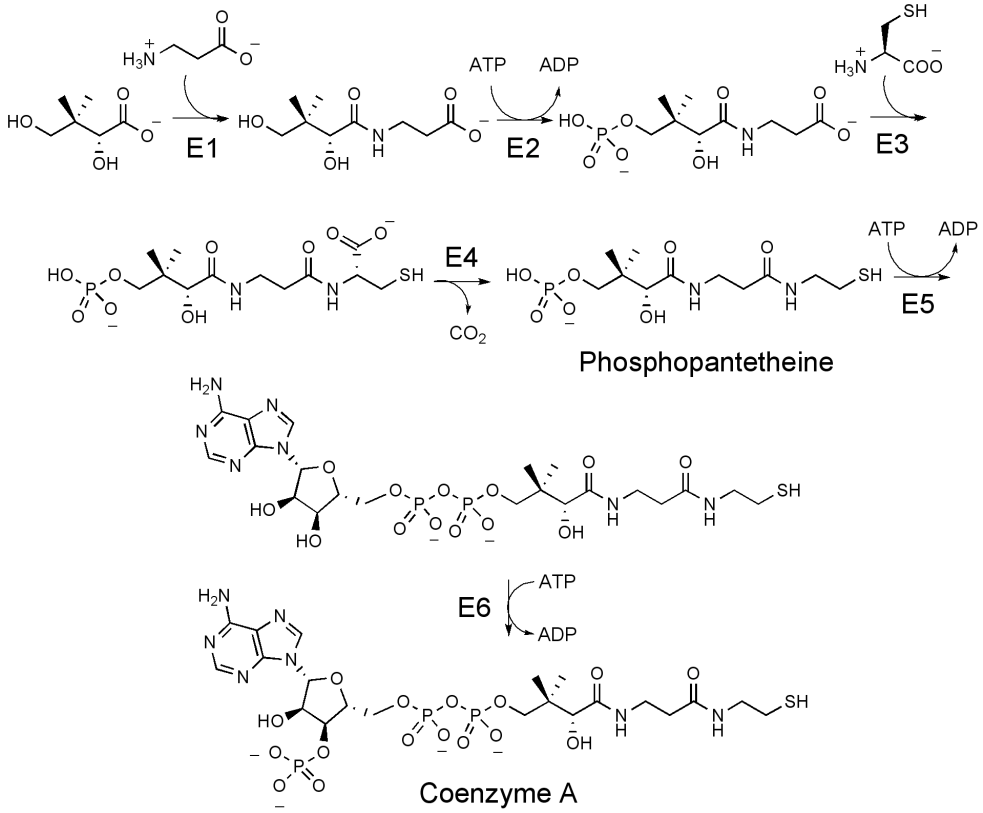


Figure 4: Biosynthesis of coenzyme A. E1 = pantoate-b-alanine ligase, E2 = pantothenate kinase, E3 = phosphopantothenoyl cysteine synthetase, E4 = phosphopantothenoyl cysteine decarboxylase, E5 = phosphopantetheine adenylyl transferase (PPAT), E6 = dephospho coenzyme A kinase (DPCK)

terminal alcohol, adenylated at that phosphate, and finally phosphorylated on the ribose ring. Our synthesis works in a similar fashion. In fact, three enzymes from CoA biosynthesis were cloned from *E. coli* genomic DNA and used for the final three steps of our synthesis.

Ye Liu developed the protocol for amino-CoA synthesis. The chemoenzymatic synthesis began with a solid-phase strategy to prepare the amino-pantetheine portion of

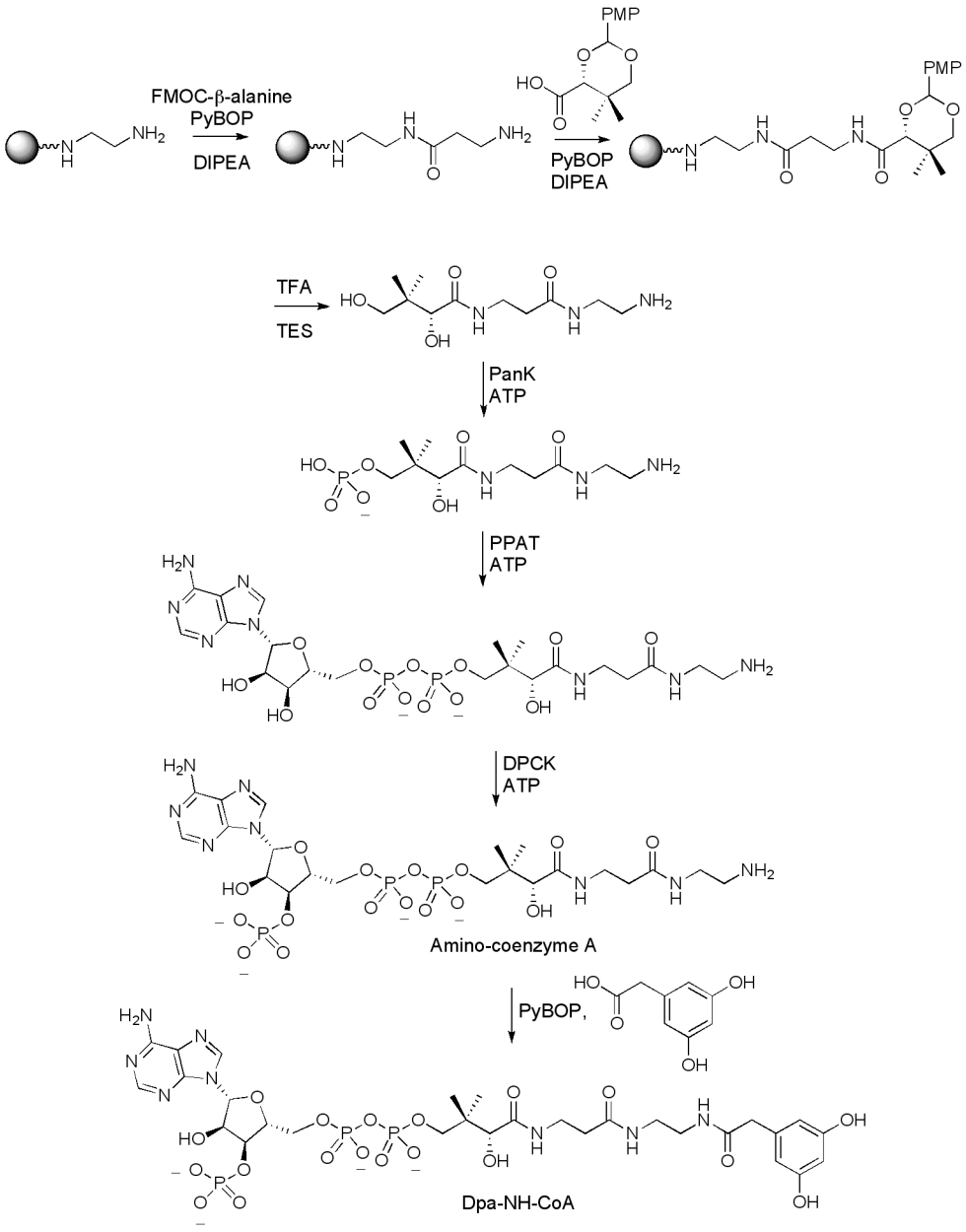


Figure 5: Chemoenzymatic synthesis of amino-coenzyme A

amino-CoA (Figure 5).[11,13] First, trityl resin was loaded with diaminoethane. FMOC-protected β-alanine was then coupled to resin-bound diaminoethane. After removal of the FMOC group, PMP-protected pantoic acid was coupled to the dipeptide. Simultaneous deprotection and release from the resin resulted in amino-pantetheine. The synthesis is completed with the enzymes: pantetheine kinase (panK), phosphopantetheine

adenylyltransferase (PPAT) and dephosphocoenzyme A kinase (DPCK) in a one-pot reaction. This synthesis is an efficient and practical route to amino-CoA analogs, which are powerful tools in mechanistic enzymology. It is estimated that 4% of all enzymatic reactions use coenzyme A as a cofactor. Amino-CoA can be used to make non-hydrolyzable analogs for an enormous variety of enzyme-catalyzed reactions. In the case of DpgC, coupling of amino-CoA to dihydroxyphenylacetic acid gives the substrate analog dihydroxyphenylacetyl-amino-CoA (Dpa-NH-CoA).

Ye Liu carried out inhibition studies of Dpa-NH-CoA with DpgC. Inhibition assays were performed using the DTNB assay, which measures the presence of free thiols through formation of a spectroscopically active disulfide.[14] The K_i of Dpa-NH-CoA was determined to be 2.7 ±0.3 μM, which is comparable to the Michaelis constant (K_M) of the natural substrate. Dpa-NH-CoA was recovered unchanged from DpgC assays. Once Dpa-NH-CoA was established as a competitive inhibitor with reasonable kinetic parameters, crystallography experiments with DpgC and inhibitor were performed.

Crystallization of DpgC with Dpa-NH-CoA

DpgC was overexpressed and purified as described in chapter 2. Dpa-NH-CoA (2 mM) was incubated with DpgC (12 mg/mL in 20 mM Tris, 100 mM NaCl at 20 ^{o}C for 2 hr). The enzyme/inhibitor mixture was screened against Hampton Research Crystal Screens I and II. Inhibitor bound DpgC crystallized in 100 mM sodium citrate, 165 mM ammonium acetate and 24% (w/v) PEG 4000, pH 5.6, after two days. The enzyme inhibitor complex crystallized in the space group $P2_12_12$ with two trimers per asymmetric unit based on the Matthews coefficient.[15]

Structure solution of DpgC bound to Dpa-NH-CoA and model building

Phase information for the structure of DpgC bound to the substrate analog was calculated with the AMORE[16] software package using the incomplete native structure, discussed in Chapter 2, as the search model. AMORE found three monomers in the asymmetric unit that made up a trimer. The biological unit of DpgC, a dimer of trimers, results from symmetry operations in the crystal. Electron density corresponding to the protein is consistent and well defined throughout all three monomers. The active site is well ordered and the substrate analog is clearly evident in the electron density maps. Unfortunately, the second trimer predicted by the Matthews coefficient could never be built. An area in the asymmetric unit that would fit a second trimer contains electron density that does not resemble a protein structure. It was suggested that this missing trimer was the result of lattice-translocation defect.[17] This rare anomaly describes a protein that occupies several positions in the unit cell. Density from such a region is not interpretable. Although lattice translocation defect could explain the phenomenon the observed in our electron density maps, the effect cannot occur in crystals with $P2_12_12$ symmetry. When the data was scaled in the lower symmetry space group $P2_1$, in which lattice translocation effect is possible, the incomplete electron density is still observed and still uncorrectable. In an attempt to fill the empty portion of the unit cell, lower occupancy, overlapping trimers were placed using search programs such as AMORE. These experiments failed to place any trimers in the unit that improved refinement statistics. After consulting with several experts in

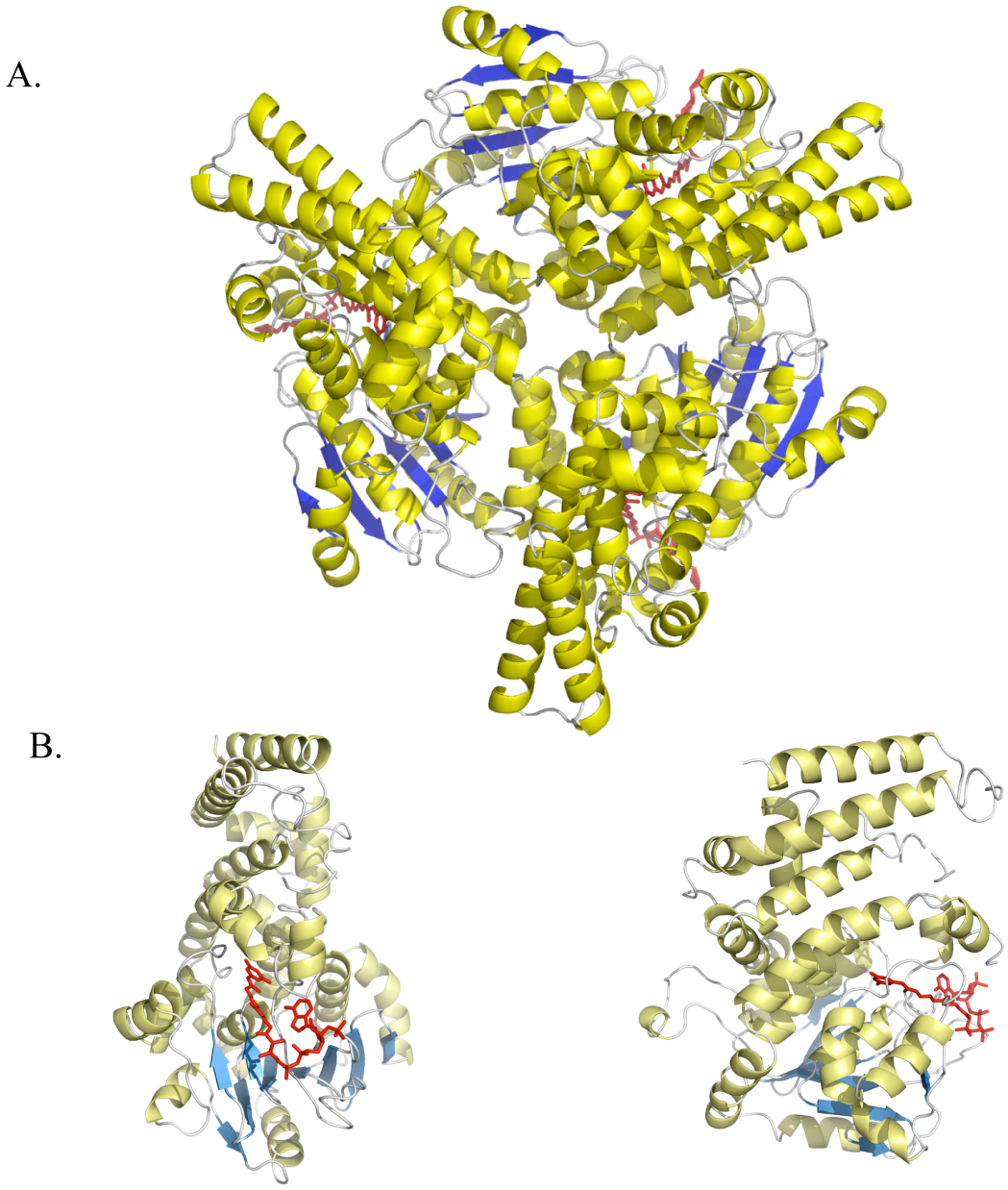


Figure 6: A. DpgC trimer looking down three-fold axis of symmetry. B. DpgC monomer, with substrate mimic colored red.

the field of crystallography, proper identification and correction of the anomaly could not be reached.[18,19,20] However, we are confident that the model of DpgC bound to the

substrate mimic is a complete and accurate picture of the enzyme in a starting material complex. All conclusions based on crystallographic evidence are corroborated with biochemical evidence. Overall, the complex of DpgC with Dpa-NH-CoA provided an exciting look at how this unusual enzyme activates substrate and molecular oxygen for downstream chemistry.

The active site of DpgC bound to Dpa-NH-CoA

As mentioned, the active of DpgC and the substrate mimic are well defined by the electron density maps. The coenzyme A portion of the substrate mimic binds in the active site in an extended conformation with the adenosine moiety bent back around the molecule. Lys238 and His222 make hydrogen bonding contacts with the phosphate group on the 4' position of the ribose ring (Figure 7). Tyr225 makes a hydrogen bonding contact with the β-phosphate of the substrate mimic. The conformation of the adenosine group over the pantetheine portion of CoA is a typical binding motif in the crotonase superfamily.[21] The N7 and exocyclic amine of the adenosine moiety make hydrogen bonds with the pantetheine group further stabilizing the observed conformation. These aspects of substrate binding are important,

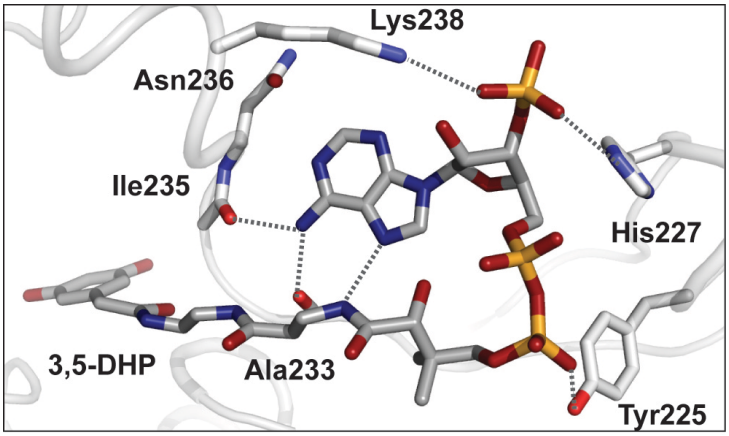


Figure 7: Dpa-NH-CoA bound to DpgC. DpgC shown in surface, substrate mimic shown in stick and molecular oxygen in sphere representation.

as an *N*-acylcysteamine derivative of the substrate lacking the adenosine and most of the pantetheine portion is not turned over.[22] However, these binding features do not reveal how DpgC achieves catalysis.

The 3,5-dihydroxyphenyl ring of the substrate mimic bound to DpgC gives a clear picture of how the substrate is activated for catalysis. The 5-hydroxyl of Dpa-NH-CoA forms a water bridge with Glu255. This is one interaction that creates a build-up of negative charge on the substrate. The 3-hydroxyl group makes a unique interaction with two residues: Glu189 forms a hydrogen bond with the 3-OH, while the face of the guanidinium group of Arg254 stabilizes the emerging negative charge. Electron density maps clearly show the face of Arg254 closest to the 3-OH. To our knowledge, this is the first observation of the ζ-carbon of an arginine residue forming such an interaction. With a negative charge being created on the 3-OH, one would expect the amine groups on Arg254 to quench the charge. However the ζ-carbon of Arg254 is closest to the 3-OH, which positions an atom

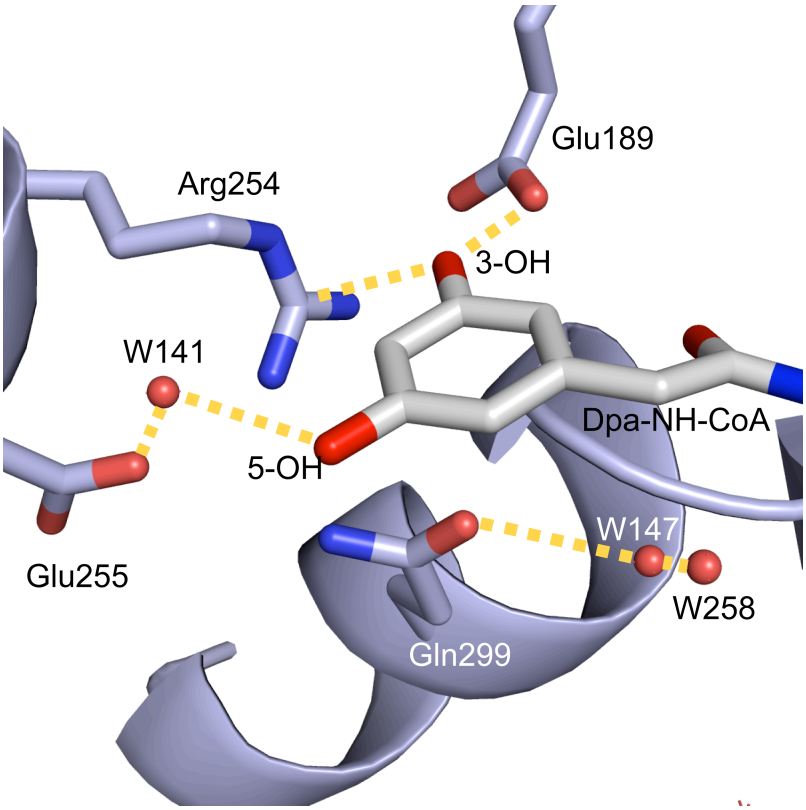


Figure 8: Recognition of the 3,5-dihydroxyphnyl ring of the substrate by DpgC

with a partial positive charge near the hydroxyl group. This is analogous to the function of histidine residues in flavin chemistry.[6] The Arg254/Glu189 interaction is critical to catalytic turnover as demonstrated by biochemical work discussed later in this chapter. The described interaction of Arg254/Glu189 with the 3-OH demonstrates a build up of negative charge around the dihydroxyphenyl ring; however there is another crucial component of DpgC/substrate binding that prepares the substrate for catalysis.

The oxyanion hole discussed in the second chapter is clearly observed in the structure bound to the substrate mimic (Figure 9). The carbonyl of the substrate mimic is ideally oriented between the amide hydrogens of Gly296 and Ile235, which stabilizes the predicted thioester enolate intermediate in the reaction pathway. Formation of this enolate results from deprotonation at the α-C position of the substrate. The oxyanion hole functions to lower the pKa of this proton significantly. A water

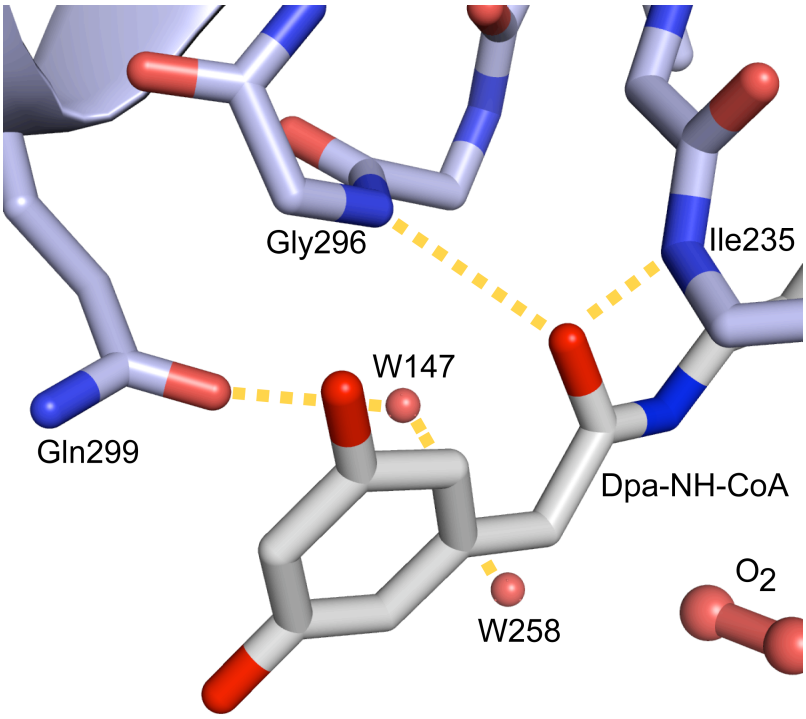


Figure 9: DpgC bound to the substrate mimic, Gly296 and Ile235 form the oxyanion hole

network anchored by Gln299 facilitates deprotonation at the α-C position. Glutamine is conserved in all DpgC homologues at this location. The use of a glutamate to anchor the water network would provide a stronger base for deprotonation, suggesting a role for Gln299 beyond deprotonation. This possibility will be discussed chapter four. Due to the location of the water network anchored by Gln299, deprotonation is predicted to occur at the *proR* hydrogen. Thus oxygen will most likely attack the opposite face of the substrate.

At this position we observed unassigned electron density consistent in size and shape with molecular oxygen (Figure 10). Each monomer of the trimer has density in the same location and similar orientation. The orphan electron density is too large to be a water molecule. The hydrophobic nature of the binding pocket makes the electron density unlikely to be a cationic metal. The crystallization media contains no compounds that would fit in the binding pocket. Based on the chemistry of the enzyme and the environment of the potential oxygen-

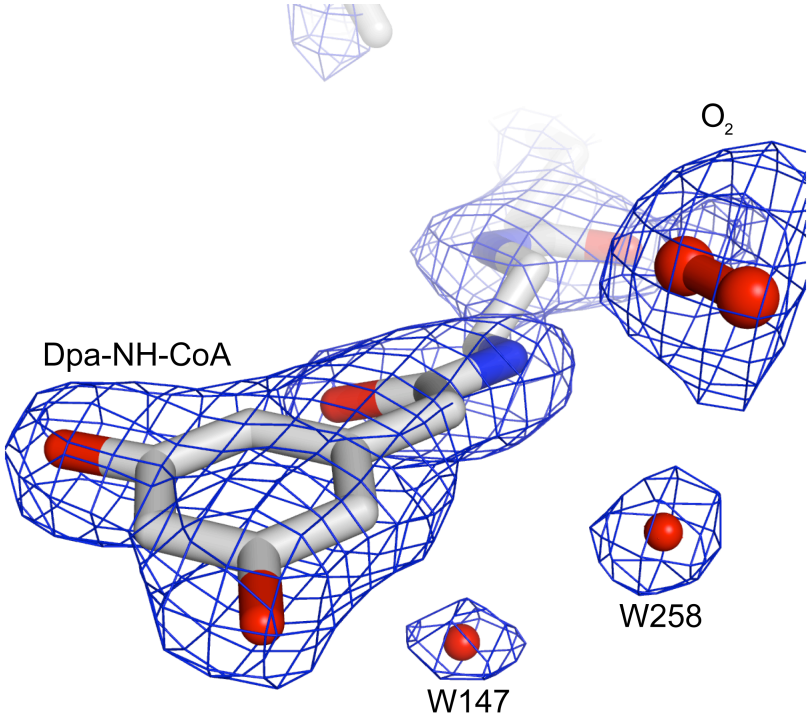


Figure 10: A composite omit electron density map showing the area around the substrate mimic, catalytically relevant waters and molecular oxygen.

binding pocket, it is reasonable to assign this electron density as molecular oxygen. Ordered molecular oxygen is rarely observed in protein crystal structures,[23,24] especially when not in complex to a transition metal.[25,26]

Unequivocal assignment of the electron density as molecular oxygen is not possible based solely on the X-ray diffraction data. Experiments regarding corroboration of the electron density as oxygen and probing the hydrophobic pocket through biochemical studies are discussed in the next chapter of this thesis. Based on the evidence discussed in this chapter, a chemical mechanism for the dioxygenation reaction of DpgC was proposed (Figure 11).

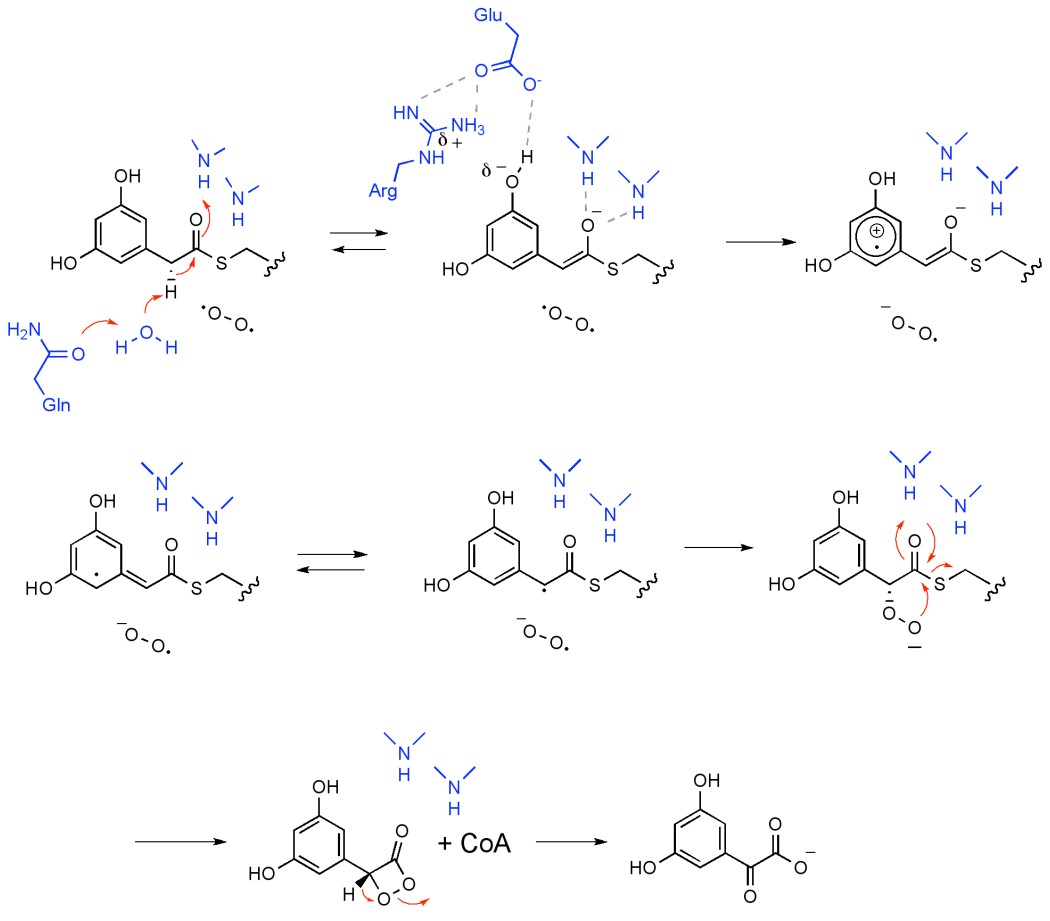


Figure 11: Proposed mechanism for the cofactor-independent dioxygenase DpgC

The first step of the proposed reaction is deprotonation at the α–C of the substrate and formation of a thioester enolate. The backbone amides of Gly296 and Ile235 stabilize this intermediate in the tradition of the crotonase superfamily.[21] The thioester enolate combined with the Arg254 and Glu189 interactions with the 3-OH to create a highly conjugated, negatively charged intermediate. Molecular oxygen is oriented in the active site to react with this intermediate. Similar to flavin chemistry, we propose a reaction mechanism with molecular oxygen is a two-step process. The first step is single electron transfer from the substrate to triplet oxygen. This reduction forms a conjugated radical cation/superoxide pair. The second step of molecular oxygen activation is reaction of the radical cation substrate intermediate with superoxide in a spin-allowed, bond-forming process to give the peroxide intermediate.

Based on modeling intermediates in the active site, Gln299 may play a role in stabilizing the peroxy intermediate. It has been suggested that a glutamine or asparagine residue performs stabilization of peroxy intermediates in flavin and cofactor independent oxygenases. Such a role may explain the use of a glutamine in this position as opposed to a glutamate. The peroxide intermediate can attack the thioester bond, still positioned in the oxyanion hole, which releases coenzyme A and forms a 1,2-dioxetanone intermediate. DpgC must carefully direct break down of the resulting dioxetanone to produce the glyoxylate

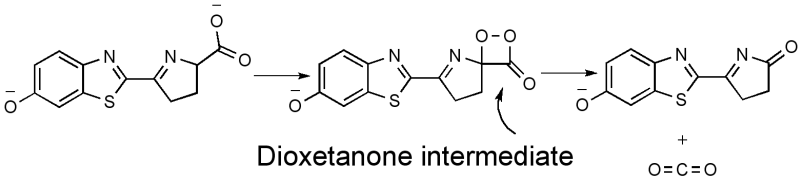


Figure 12: The reaction catalyzed by firefly luciferase,which stabilizes a 1,2-dioxetanone intermediate, the final product is in an excited state and emits light

product and avoid decomposition to carbon dioxide. The 1,2-dioxetanone species is observed in firefly luciferase (Figure 12).[27] In this system the dioxetanone breaks down to CO_2 and product. There is also biochemical evidence for the organic radical and superoxide pair in the cofactor independent dioxygenase Hod.[7] The suggested role of Arg254 in DpgC, which stabilizes a negative charge on the 3-OH, is analogous to the proposed role of a histidine residue in stabilizing reduced flavin cofactors. The role of Arg254 as well as features of substrate recognition was further explored through crystallographic and biochemical characterization.

Mutagenesis studies of substrate recognition and catalysis by DpgC

Elisha Fielding performed mutation studies of several active site residues. These studies showed residues Arg254 and Glu189 to be most important to catalysis as the Arg254Lys and Glu189Gln constructs showed significant decreases in enzyme kinetics. These results emphasize the importance of this unique dyad in catalysis of the enzyme. Two mutant constructs of active site residues, Glu255Gln and Gln299Asn, showed comparable kinetic statistics to the wild type.[22]

Table 1: Kinetic parametes for DpgC mutant enzymes with the natural substrate Dpa-CoA

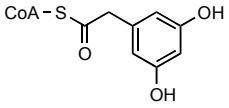
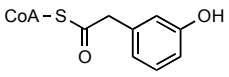
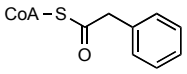
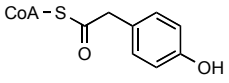
DpgC	K_M **(μM)**	k_{cat} **(min^{-1})**	k_{cat}/K_M
wild type	3.9 ±0.6	10.32 ±0.42	1
R254K	217 ±46	5.17 ±0.32	0.009
E189Q	64 ±13	5.12 ±0.41	0.029
E255Q	3.6 ±1.9	9.18 ±0.78	0.959
Q299N	2.5 ±1.1	4.14 ±0.21	0.619

Kinetic parameters, K_M and k_{cat}, of alternate substrates were also determined in order to establish the catalytic significance of the hydroxyl groups on the phenyl ring of

the substrate (Figure 13). 3-hydroxyphenylacetyl CoA has kinetic parameters comparable to the natural substrate. The other two alternate substrates showed significantly decreased turnover and binding affinity.

The data from the R254K and E189Q mutants along with the alternate substrates indicates the importance of a single hydroxyl group forming a phenoxide on the ring. The role of the atypical interaction between Arg254 and the 3-OH could be necessary to stabilize the negative charge on the 3-position phenoxide without presenting an acidic proton.

Table 2: Kinetic parameters of alternate Phenylacetyl-CoA substrates for DpgC

Substrate	K_M **(μM)**	k_{cat} **(min^{-1})**	k_{cat}/K_M*
	3.9 ±0.6	10.32 ±0.36	1
	4.4 ±1.4	7.52 ±0.57	0.650
	102 ±37	3.84 ±0.36	0.014
	851 ±361	13.32 ±2.33	0.006

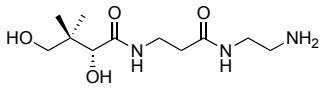
Conclusion

The structure of DpgC bound to a substrate mimic has elucidated the chemistry of this unusual enzyme. The identification of an oxygen binding hydrophobic pocket is a key discovery in the field of enzymatic oxygen activation. The structure reveals novel

insights that can be applied to the general mechanism of oxygen activation. Future work will involve crystallographically characterizing intermediates along the DpgC pathway.

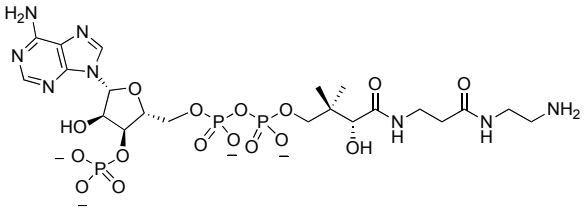
Materials and methods

Synthesis of dihydroxyphenylacetyl-amino-CoA (Dpa-NH-CoA)[11]



Synthesis of amino-pantetheine. PMP-protected pantoic acid was synthesized using the published procedure.[13] Panetheine analog was synthesized on solid support using trityl chloride polystyrene resin (Novabiochem, loading 1.6 mmol/g) in a solid phase synthesis vessel. The resin (1.00g, 1.6mmol) was swelled in DCM (8 mL) for 1 hr. before 1,2 diaminoethane (1.07mL, 16.0 mmol) was added. The suspension was agitated on a shaker for 16 hr., then drained and rinsed with NMP (*N*-methyl-2-pyrrolidinone) 5x20 mL to give 1,2-diaminoethane-functionalized resin. A solution of FMOC-β-alanine (1.49 g, 4.80 mmol) PyBOP (2.50g , 4.80 mmol), and DIPEA (1.67 mL, 9.60 mmol) in NMP (4 mL) was added to the resin and the suspension was agitated for 2 hr., then drained and rinsed with NMP (5x20 mL). The FMOC protecting group was removed by treating the resin with 20% (v/v) piperidine/NMP (20 mL) for 20 min, which was then rinsed with NMP (5x 20 mL). A solution of PMP-protected pantoic acid (1.28 g, 4.80 mmol), PyBOP (2.50 g, 4.80 mmol) and DIPEA (1.67 mL, 9.60 mmol) in NMP (4 mL) was added to the resin and the suspension was agitated for 4 hr. After rinsing with NMP (5x 20 mL) and DCM (5x 20 mL), the resin was treated with 15 mL of 5% (v/v) trifluoroacetic acid (TFA) and 2% (v/v) triethylsilane (TES) in DCM for 20 min to cleave the product from the resin. The resin was washed with 5% TFA in DCM (10 mL). The cleavage solution and wash solution were combined and concentrated *in vacuo*. The residue was redissolved in 10

mL of 10% (v/v) TFA/H_2O. The aqueous solution was extracted twice with diethyl ether (10 mL), and then lyophilized to give amino-pantetheine as a yellow oil. The product was further purified by preparative HPLC (9 mL/min flow rate; gradient: 0-5 min, 5% B; 5-25 min, 5-98% B, where A=0.1% TFA/H_2O and B=CH_3CN). Chromatographs were monitored at 220 nm and amino-pantetheine eluted at 11.2 min. Removal of solvents via lyophilization gave amino-pantetheine as a colorless oil (combined purified yield 70%, 290 mg). **^{1}H NMR** (D_2O, 300 MHz) δ 3.94 (s, 1H), 3.49-3.43 (m, 5H), 3.35 (d, J = 11.1 Hz, 1H), 3.10 (t, J = 5.9 Hz, 2H), 2.48 (t, J = 5.9 Hz, 2H), 2.48 (t, J = 6.5 Hz, 2H), 0.88 (s, 3H), 0.85 (s, 3H). **^{13}C NMR** (D_2O, 100 MHz) δ 175.1, 147.9, 75.9, 68.5, 39.4, 38.8, 37.0, 35.6, 20.7, 19.4. **HRMS** (ESI+) *m/z* Calcd for $C_{11}H_{24}N_3O_4$ 262.1767, found 262.1768.



Enzymatic preparation of amino-coenzyme A. The genes encoding pantetheine kinase (PanK), phosphopantetheine adenylyltransferase (PPAT) and dephospho coenzyme A kinase (DPCK) were each amplified from *E. coli* genomic DNA using the primers shown below:[28]

PanK:

5’ 5’-GGGAATTCCATATGACCGCCAGAAACATGCTTATGAG

3’ 5’-CGCGGATCCAAGCTTTTATTTGCGTAGTCTGACCTCTTCTACCG

PPAT:

5’ 5’-GTCTCTAGAGCTAGCATGCAAAAACGGGCGATTTATCC

3’ 5’-CGCGGATCCAAGCTTCTACGCTAACTTCGCC

DPCK:

5’ 5’-GGGAATTCCATATGAGGTATATAGTTGCCTTAACGGGAG

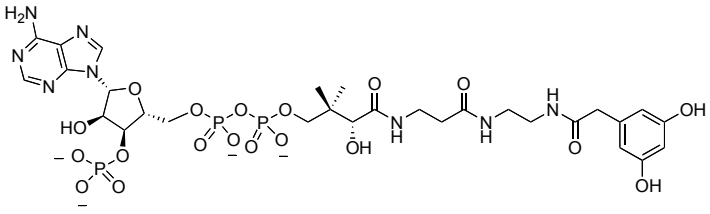
3’ 5’-CGCGGATCCAAGCTTTTACGGTTTTTCCTGTGAGACAAACTGC

The amplified genes were digested (*Nde*I and *Hind*III for PanK and DPCK, and *Nhe*I and *Hind*III for PPAT) and ligated into the pET28a vector (Novagen). The vectors were individually transformed and proteins were overexpressed in *E. coli* BL21(DE3) cells (Novagen). For each enzyme, cells were harvested by centrifugation at 3500 rpm for 20 min. The cell pellets were suspended in 40 mL lysis buffer (20 mM Tris, 500 mM NaCl, pH 7.5) and lysed by passage through a French Press cell at 1000 psi. The lysate was centrifuged at 10,000 rpm for 20 min and the supernatant was incubated with 1 mL Ni-NTA resin (Qiagen) for 1 hr. The resin was washed with lysis buffer (4x 15 mL) and the enzyme was eluted with 2x 15 mL elution buffer (20 mM Tris, 500 mM NaCl, 250 mM imidazole, pH 7.5). The elution fractions were dialyzed against 1 L of 50 mM HEPES, 250 mM NaCl, 2 mM $MgCl_2$, pH 8.0 for 2 hr at 4^{o}C.

Enzymatic synthesis reactions were carried out at 37^{o}C in a single-pot reaction. 39.3 mg of amino-pantetheine was added to 30 mL of buffer solution (20 mM KCl, 10 mM $MgCl_2$, and 50 mM Tris, pH 9.0). Next, 10 mg of PanK along with 152 mg of ATP was added and the reaction was allowed to shake for 30 min. 10 mg of PPAT in addition to 152 mg of ATP was added and the reaction was allowed to shake for 30 min. Finally, 10 mg of DPCK with 152 mg of ATP was added and the reaction shook for 45 min. To terminate the reaction, 50% w/v trichloroacetic acid (TCA) was added to give a final concentration of 10% w/v in the reaction mixture. After incubation on ice for 10 min, the

mixture was centrifuged at 5,000 rpm for 10 min. The supernatant was lyophilized, and the crude reaction mixture was purified by preparative reverse-phase C18 HPLC using the following gradient: 9 mL/min; 0-5 min, 0% B; 5-25 min, 0-5% B, where A=0.1% TFA/H_2O and B=CH_3CN; amino-CoA eluted at 25.5 min. After lyophilizing the purified product yield was 50%.

Amino-coenzyme A Analog. **^{1}H NMR** (D_2O, 400 MHz) δ 8.69 (s, 1H), 8.45 (s, 1H), 6.24 (d, *J* = 5.5 Hz, 1H), 4.89-4.87 (m, 2H), 4.62 (br s, 1H), 4.29-4.27 (m, 2H), 4.04 (2, 1H), 3.84 (q, *J* = 5.0 Hz, 1H), 3.66 (q, *J* = 6.0 Hz, 2H), 2.51 (t, *J* = 6.5 Hz, 2H), 0.95 (s, 3H), 0.89 (s, 3H). **^{13}C NMR** (D_2O, 126 MHz) δ 175.4, 175.0, 150.1, 148.7, 144.9, 142.7, 118.8, 87.7, 83.8, 74.7, 74.3, 74.2, 71.9, 65.3, 39.4, 38.6, 37.0, 35.8, 35.6, 20.8, 19.1. **^{31}P NMR** (D_2O, 121 MHz) δ 0.81, -9.87, -10.30. **HRMS** (ESI+) *m/z* Calcd for $C_{21}H_{38}N_8O_{16}P_3$ 751.1619, found 751.1653.



Synthesis of Dpa-NH-CoA. The sustrate mimic was synthesized by coupling 3,5-dihydroxy-phenylacetic acid with amino-CoA. To prepare 3,5-dihydroxy-phenylacetic acid, methyl 3,5-dihydroxyphenylacetate (150 mg, 0.823 mmol) was added to 5.0 mL of 2 M aqueous NaOH and stirred at room temperature. After gentle shaking for 1.5 hr, the reaction was acidified to pH 1.0 with 2 M HCl. The resulting orange solution was flash frozen and lyophilized. A red-orange powder was obtained and taken up in MeOH and passed through a plug of silica gel. Concentration *in vacuo* gave 3,5-dihydroxy-phenylaceitc acid as a white poweder (110 mg, yield 80%).

The coupling of 3,5-dihydroxy-phenylacetic acid and amino-CoA was done following the published procedure for the synthesis of acyl amino acid coenzyme A thioester,[29] with two modification: 2.5 eq. of 3,5-dihydroxy-phenylacetic acid and 2.5 eq. of PyBOP were used to couple 9.0 mg of amino-CoA. The crude product was purified by preparative HPLC using the following gradient: 8 mL/min; 0-3 min, 0% B; 3-40 min 0-20% B, where A=0.1% TFA/H_2O; B=CH_3CN; Dpa-NH-CoA eluted at 222 min, monitoring at 260 nm. After lyophilizing, Dpa-NH-CoA was recovered as a white powder (10.3 mg, yield 95%). **^{1}H NMR** (D_2O, 400 MHz) δ 8.71 (s, 1H), 8.45 (2, 1H), 6.31 (overlapping m, 2H), 6.24 (overlapping m , 2H), 4.93 (overlapping m, 2H), 4.66 (br s, 1H), 4.34 (br s, 2H), 4.07 (s, 1H), 3.92-3.90 (m, 1H), 3.69-3.67 (m, 1H), 3.44 (overlapping m, 4H), 3.33 (overlapping m, 4H), 2,42 (t, *J* = 6.0 Hz, 2H), 0.99 (s, 3H), 0.89 (s, 3H). **^{13}C NMR** (D_2O, 126 MHz) δ 174.9, 174.8, 174.3, 157.1 150.0, 148.6, 144.8, 142.7, 137.7, 118.8, 108.3, 101.5, 87.8, 83.8, 74.6, 74.3, 74.2, 72.1, 65.3, 42.5, 38.9, 38.8, 38.6, 35.6, 21.0, 18.8. **^{31}P NMR** (D_2O, 121 MHz) δ 0.74, -9.90, -10.34. **HRMS** (ESI-) *m/z* Calcd for $C_{29}H_{42}N_8O_{19}P_3$ 899.1779, found 899.1825.

Crystallographic studies of DpgC with the substrate mimic Dpa-NH-CoA[15]

Crystallization of DpgC with Dpa-NH-CoA. Elisha Fielding established crystallization conditions for DpgC bound to Dpa-NH-CoA. DpgC (12 mg/mL) was incubated with Dpa-NH-CoA (2 mM) at 20^{o}C for 2 hr. Crystallization conditions were determined by screening the enzyme small molecule mixture against the Hampton Research Crystal Screens I and II (Hampton). The co-complex crystallized in 100 mM sodium citrate, 165

mM ammonium acetate and 24% (w/v) PEG 4000, pH 5.6. The DpgC/substrate mimic complex (1.5 μL) was mixed with 1.5 μL of the crystallization solution. Crystals typically appeared after 2 days. Crystals were transferred to a cryoprotectant solution, consisting of crystallization solution with 20% glycerol, then were flash frozen in liquid N_2.

Structure determination. Phase information for the co-complex structure was calculated with the molecular replacement program AMORE using the incomplete native structure as the search model. The model was built with the program COOT[30] and refined with the software suite CNS.[31] Initial structural refinement was performed with non-crystallographic (NCS) restraints. After several rounds, the restraints were removed from calculations. Sigma-weighted simulated annealing composite omit maps were used to build the model. The Matthews coefficient indicated two trimers present in the asymmetric unit. Due to an anomaly in the crystal, only one trimer was completed. This anomaly is discussed earlier in the chapter.

Table 3: Crystallography data collect and refinement statistics

Data collection	DpgC/Dpa-NH-CoA complex
Space group	$P2_12_12$
Cell dimensions	
a, b, c (Å)	139.9, 156.7, 171.0
α, β, γ, (°)	90, 90, 90
Wavelength	1.000001
Resolution (Å)	2.40
R_{sym} or R_{merge}	0.130 (0.577)
$I/\sigma I$	15.46 (2.55)
Completeness (%)	98.0 (96.7)
Redundancy	5.1 (4.4)
Refinement	
Resolution (Å)	2.45
No. reflections	128559
R_{work}/ R_{free}	0.328/0.356
No. atoms	
Protein	9797
Ligand/ion	180
Water	225
B-factors	
Protein	38.82
Ligand/ion	42.84

Water	32.83
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (^{o})	1.4827

*Highest resolution shell is shown in paranthesis.

Kinetic studies of DpgC mutants and alternate substrates. Ye Liu determined kinetic parameters of Dpa-NH-CoA.[11] Elisha Fielding carried out all kinetic studies on DpgC mutants Arg254Lys, Glu189Gln, Glu255Gln and Gln299Asn; as well as all DpgC alternate substrate assays.[22]

References:

1 Fetzner, S., Oxygenases without requirement for cofactors or metal ions. *Appl Microbiol Biotechnol* **60** (3), 243 (2002).

2 *The Organic Chemistry of Enzyme-Catalyzed Reactions*. (Academic Press, San Diego, 2002).

3 Bugg, T. D., Oxygenases: mechanisms and structural motifs for O(2) activation. *Curr Opin Chem Biol* **5** (5), 550 (2001).

4 Wolfe, M. D., Parales, J. V., Gibson, D. T., and Lipscomb, J. D., Single turnover chemistry and regulation of O_2 activation by the oxygenase component of naphthalene 1,2-dioxygenase. *J Biol Chem* **276** (3), 1945 (2001).

5 Warshel, A., Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J Biol Chem* **273** (42), 27035 (1998).

6 Roth, J. P. and Klinman, J. P., Catalysis of electron transfer during activation of O_2 by the flavoprotein glucose oxidase. *Proc Natl Acad Sci U S A* **100** (1), 62 (2003).

7 Frerichs-Deeken, U. et al., Dioxygenases without requirement for cofactors and their chemical model reaction: compulsory order ternary complex mechanism of 1H-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase involving general base catalysis by histidine 251 and single-electron oxidation of the substrate dianion. *Biochemistry* **43** (45), 14485 (2004).

8 Gunther, M. R., Peters, J. A., and Sivaneri, M. K., Histidinyl radical formation in the self-peroxidation reaction of bovine copper-zinc superoxide dismutase. *J Biol Chem* **277** (11), 9160 (2002).

9 Stubbe, J. and van Der Donk, W. A., Protein Radicals in Enzyme Catalysis. *Chem Rev* **98** (2), 705 (1998).

10 Tseng, C. C., Vaillancourt, F. H., Bruner, S. D., and Walsh, C. T., DpgC is a metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase in the vancomycin biosynthetic Pathway. *Chem Biol* **11** (9), 1195 (2004).

11 Liu, Y. and Bruner, S. D., Rational manipulation of carrier-domain geometry in nonribosomal peptide synthetases. *Chembiochem* **8** (6), 617 (2007).

12 Mishra, P. K. and Drueckhammer, D. G., Coenzyme A Analogues and Derivatives: Synthesis and Applications as Mechanistic Probes of Coenzyme A Ester-Utilizing Enzymes. *Chem Rev* **100** (9), 3283 (2000).

13 Mandel, A. L., La Clair, J. J., and Burkart, M. D., Modular synthesis of pantetheine and phosphopantetheine. *Org Lett* **6** (26), 4801 (2004).

14 Chen, H., Tseng, C. C., Hubbard, B. K., and Walsh, C. T., Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc Natl Acad Sci U S A* **98** (26), 14901 (2001).

15 Widboom, P. F., Fielding, E. N., Liu, Y., and Bruner, S. D., Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis. *Nature* **447** (7142), 342 (2007).

16 COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4., The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D* **50**, 760 (1994).

17 Wang, J., Kamtekar, S., Berman, A. J., and Steitz, T. A., Correction of X-ray intensities from single crystals containing lattice-translocation defects. *Acta Crystallogr D Biol Crystallogr* **61** (Pt 1), 67 (2005).

18 Dauter, Zbigniew.

19 Board, CCP4 Bulletin.

20 Heroux, Annie.

21 Holden, H. M., Benning, M. M., Haller, T., and Gerlt, J. A., The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme a thioesters. *Acc Chem Res* **34** (2), 145 (2001).

22 Fielding, E. N., Widboom, P. F., and Bruner, S. D., Substrate recognition and catalysis by the cofactor-independent dioxygenase DpgC. *Biochemistry* **46** (49), 13994 (2007).

23 Taga, M. E. et al., BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature* **446** (7134), 449 (2007).

24 Eswaramoorthy, S., Bonanno, J. B., Burley, S. K., and Swaminathan, S., Mechanism of action of a flavin-containing monooxygenase. *Proc Natl Acad Sci U S A* **103** (26), 9832 (2006).

25 Berglund, G. I. et al., The catalytic pathway of horseradish peroxidase at high resolution. *Nature* **417** (6887), 463 (2002).

26 Karlsson, A. et al., Crystal structure of naphthalene dioxygenase: side-on binding of dioxygen to iron. *Science* **299** (5609), 1039 (2003).

27 Koo, J. A., Schmidt, S. P., and Schuster, G. B., Bioluminescence of the firefly: key steps in the formation of the electronically excited state for model systems. *Proc Natl Acad Sci U S A* **75** (1), 30 (1978).

28 Nazi, I., Koteva, K. P., and Wright, G. D., One-pot chemoenzymatic preparation of coenzyme A analogues. *Anal Biochem* **324** (1), 100 (2004).

29 Belshaw, P. J., Walsh, C. T., and Stachelhaus, T., Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284** (5413), 486 (1999).

30 Emsley, P. and Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60** (Pt 12 Pt 1), 2126 (2004).

31 Brunger, A. T. et al., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54** (Pt 5), 905 (1998).

Chapter 4: Probing the oxygen-binding pocket of DpgC.

Introduction

The way in which cofactor-independent oxygenases bind molecular oxygen has been an area of speculation for quite some time.[1] The presence of a hydrophobic pocket in or adjacent to the active site has been suggested as a structural motif for a cofactor-independent oxygenase to bind molecular oxygen for catalysis. Alternatively, it has also been suggested that molecular oxygen in solution could come into close enough proximity to the substrate for activation, negating the requirement for a specific oxygen-binding region.[1]

The enzyme DpgC from the vancomycin biosynthetic pathway is a cofactor-independent dioxygenase, which uses a hydrophobic pocket to bind molecular oxygen. Our structural characterization of DpgC showed electron density consistent in size and shape to molecular oxygen in a hydrophobic pocket adjacent to the active site of the enzyme.[2] DpgC uses this hydrophobic pocket to bind molecular oxygen and catalyze a four electron oxidation and thioester cleavage.[3]

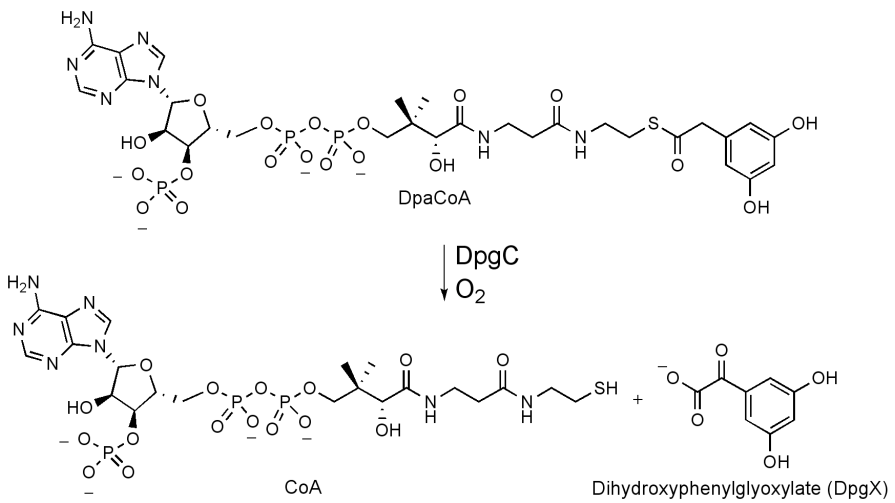


Figure 1: Reaction catalyzed by DpgC

In the case of several metal-containing enzymes, a hydrophobic pocket off the metal center has been suggested as responsible for initial oxygen binding. This hydrophobic pocket has been discussed in the examples of bovine serum amine oxidase and tyrosine hydroxylase.[4,5] Crystallographic evidence for an oxygen-binding region in flavin dependent enzymes has also been demonstrated.[6]

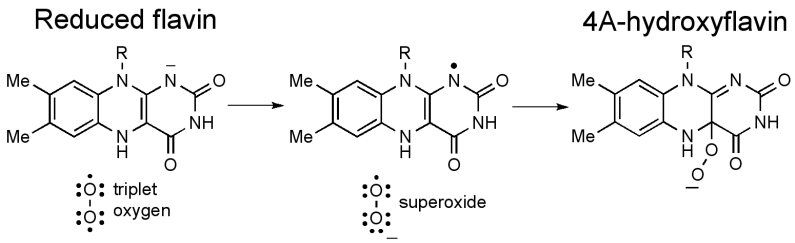


Figure 2: Activated flavin species used by flavin monooxygenase

Two recent crystal structures of flavin-dependent enzymes with oxygen bound adjacent to the flavin cofactor have been reported. Eswaramoorthy et al. solved the structure of a flavin containing monooxygenase (FMO) from *S. pombe*.[7] The authors describe an "electron density feature" present in the active site that is consistent with a bound oxygen molecule. The location of the proposed molecular oxygen molecule makes it available to form a 4A-hydroxyflavin species (Figure 2). Although molecular oxygen is a non-polar species, the authors suggest that it that forms a hydrogen bonding contact with Asn91. The presence of molecular oxygen in the crystal structure led the authors to

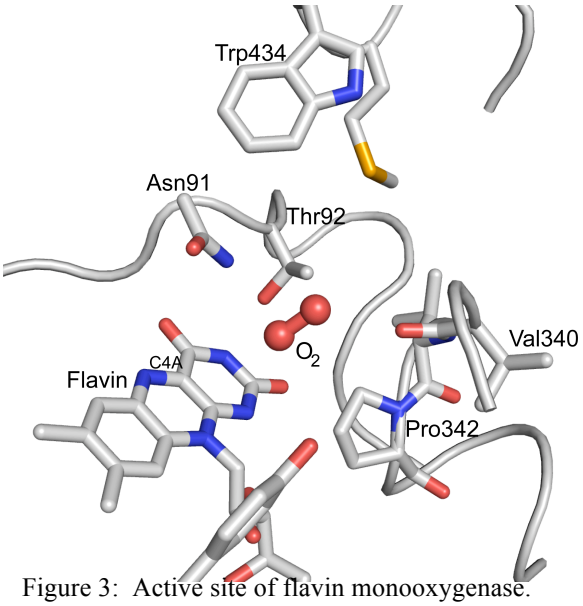


Figure 3: Active site of flavin monooxygenase.
PDB code: 1VQW

propose that Asn91 orients oxygen for reaction with flavin and may stabilize the peroxyanion intermediate.

The area in which molecular oxygen is bound in the FMO structure is not particularly hydrophobic. Aside from Asn91, residues in proximity to the modeled O_2 molecule are Thr92, Trp434, Pro342, and the carbonyl oxygen of Val340. The predicted oxygen-binding region is widely accessible to solvent. The closest contacts the oxygen molecule makes are between ND2 of Asn91, 3.26 Å to atom O1, C4A of flavin and O1, 3.57 Å, and the carbonyl of Val340 and O2 of molecular oxygen, 3.26 Å. Eswaramoorthy *et al*. suggest Asn91 is the most catalytically relevant residue as few other residues could be involved in binding and stabilizing molecular oxygen for downstream chemistry. The flavin cofactor plays the largest role in oxygen binding by providing a fairly hydrophobic environment.

The second example of oxygen bound to a flavin dependent enzyme comes from vitamin B_{12} biosynthesis. The enzyme BluB catalyzes the conversion of flavin mononucleotide to 5,6-dimethylbenzimidazole (Figure 4). The Walker lab solved the structure of BluB bound to reduced flavin.[6] The use of flavin as a cofactor and substrate is unusual and prompted the authors to name a new family after this enzyme: "flavin destructase." The authors observe electron density consistent with molecular oxygen located over the reduced flavin. As in the case of DpgC bound to the substrate mimic Dpa-NH-CoA, molecular oxygen is bound to the enzyme poised to react. Unlike DpgC, the oxygen bound to BluB is tightly held in a

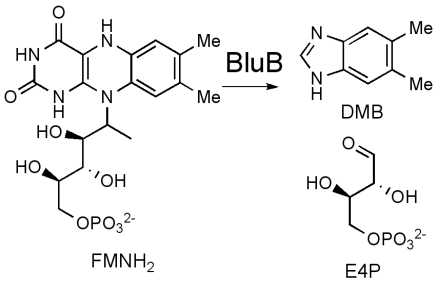


Figure 4: Chemistry of the "flavin destructase" BluB

pocket inaccessible to the solvent (Figure 4). Oxygen in the BluB structure makes two hydrogen bonds with the O2' hydroxyl of FMN and with the backbone amide of Gly61. This interaction is suggested to be a "peroxyanion hole." The authors suggest that this structural motif is shared by flavin monooxygenases as a way of binding molecular oxygen, and possibly stabilizing a peroxy-intermediate. The peroxyanion hole in BluB is

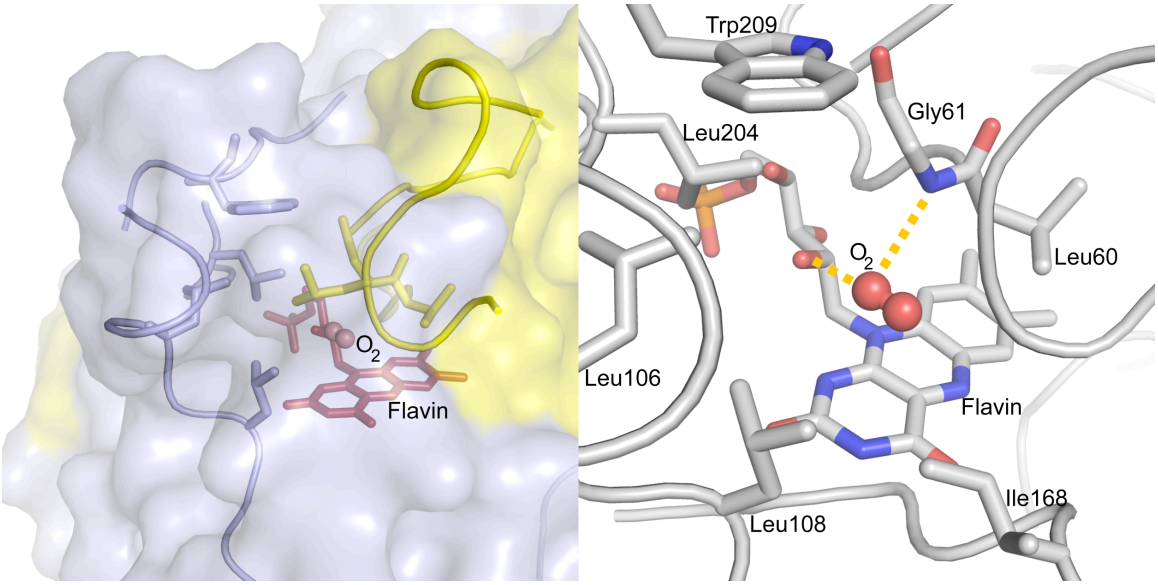


Figure 5: The active site of BluB is at the interface between two monomers. On the left, a surface representation depicting molecular oxygen and flavin buried between two monomers (monomer A in yellow, monomer B in blue). A view of the active site showing the peroxyanion hole and several residues responsible for creating the tight, hydrophobic pocket. PDB code: 2ISL

dependent on the conformation of the ribityl tail of FMN.[8,9] In the case of BluB, the ribityl tail conformation resembles that of flavin oxidoreductases not monooxygenases. BluB has conformational similarity to oxidoreductases and functional similarity to monooxygenases.

The oxygen-binding region of BluB is quite different from that of the FMO (Figure 3). BluB provides a richly hydrophobic environment with very little access to solvent. The binding pocket is formed by two monomers of BluB (Figure 5). Residues

Ile168, Leu204, Leu106, Trp209 and Leu108 from one monomer and Leu132 from the second monomer create this tight binding pocket in which FMN cradles O_2. The suggested peroxyanion hole is the only component of the pocket that is hydrophilic. The crystals of BluB with reduced flavin disintegrate over time; implying that catalysis causes a structural rearrangement that disrupts the crystal lattice. A significant conformational change would have to occur for FMN to bind the enzyme as the catalytically relevant conformation would not allow diffusion of the flavin cofactor into the active site.

As will be discussed in this chapter, the oxygen-binding pocket of DpgC shares aspects of the FMO and BluB binding pockets. Similar to BluB, DpgC has a hydrophobic pocket believed to bind molecular oxygen. However, the pocket is readily accessible to solvent as seen with FMO. We have shown experimentally that perturbing the oxygen-binding pocket through mutation and biochemical characterization has demonstrated the importance of the hydrophobic residues to oxygen binding. Crystallography with a slightly altered substrate mimic demonstrated the sensitivity of the pocket to subtle changes in the environment.

Identification of residues composing the hydrophobic pocket

The hydrophobic oxygen-binding pocket of DpgC resembles a basket that is open to solvent.[2] As oxygen moves towards the site of oxidation on the substrate, the pocket narrows and becomes capable of binding and orienting molecular oxygen for activation by the substrate. Based on the location of the electron density that is consistent with molecular oxygen, four residues were selected for mutagenesis to probe the oxygen-

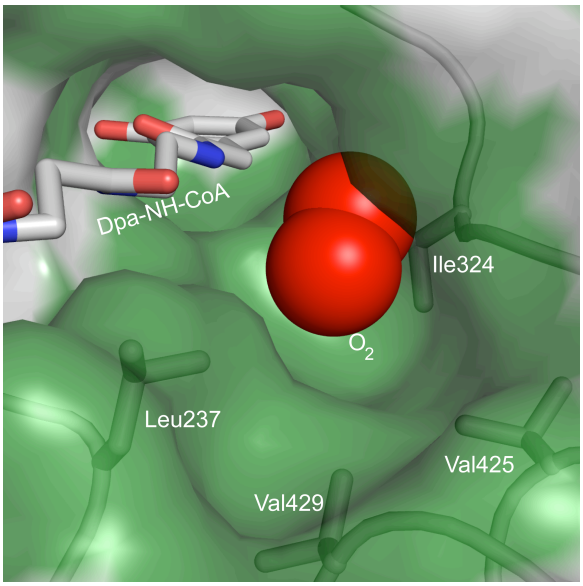


Figure 6: Surface representation of the DpgC oxygen-binding pocket. Hydrophobic residues shown in green, molecular oxygen in red

binding pocket. Although these four residues do not comprise the entire oxygen-binding pocket, they are the closest in proximity to the modeled O_2 molecule. Residues Ile324, Leu237, Val425 and Val429 were each mutated to a threonine to study the effect of subtle perturbation on the oxygen-binding pocket.

Site-directed mutagenesis in the oxygen-binding pocket and kinetic analysis of mutant constructs

Threonine was selected as the mutant for each of the four residues identified as critical in the oxygen-binding pocket. The choice of threonine provided a hydrophilic component to the hydrophobic pocket while maintaining some hydrophobicity. This change creates a subtle perturbation of the hydrophobic pocket. We felt substituting serine or aspartate residues would too drastically affect substrate binding and cause too severe a change in the active site to offer any meaningful observations. Mutant constructs were created using the QuikChange mutagenesis kit (Strategene) and were assayed for kinetic competence. Kinetic

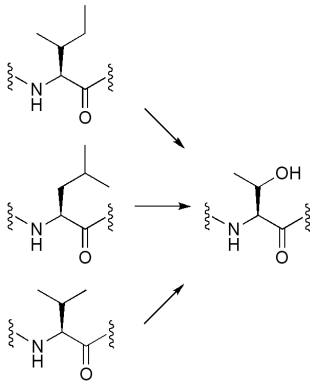


Figure 7: Mutation strategy for perturbing the DpgC oxygen-binding pocket. Mutating hydrophobic residues to threonine.

parameters, k_{cat} and K_M, of the hydrophobic pocket mutants were initially determined for the natural substrate of DpgC using the DTNB assay.[10] However, to better understand the capability of the oxygen-binding pocket, assays were carried out on an oxygen electrode. By using an oxygen electrode, we observed consumption of molecular oxygen by DpgC during catalysis. The rate of oxygen consumption was measured at various O_2 concentrations. The I324T mutant had the most significant impact on substrate as well as oxygen kinetic parameters. This was expected as Ile324 is closest to molecular oxygen in the crystal structure, 3.24 Å. In fact, this mutant showed no turnover during kinetic assays with regard to oxygen. The other three mutants also showed decreased kinetic competence. This biochemical characterization corroborated our identification of the unassigned electron density as molecular oxygen. Our kinetic data also indicated that we had identified an oxygen-binding pocket in a cofactor-independent oxygenase.

Table 1: Kinetic parameters for the DpgC hydrophobic pocket mutants

	K_M (DpaCoA)	k_{cat} (DpaCoA)	k_{cat}/K_M*	K_M (O_2)
Wild type	3.9	0.172	1.0	1.7
Ile324Thr	>1,000	ND	ND	ND
Leu237Thr	58.0	0.050	0.019	2.1
Val429Thr	14.7	0.019	0.029	>2.6
Val425Thr	13.8	0.037	0.060	2.6

The units for K_M(DpaCoA) are μM, k_{cat} are in s^{-1} and K_M (O_2) are mM. ND, not determined. * Apparent k_{cat}/K_m relative to the wild type enzyme

Soaking co-crystals with the noble gas xenon

Further attempts to unequivocally assign the orphan electron density as molecular oxygen involved a unique crystallographic experiment. The noble gas xenon (Xe) has been exploited as an isoelectronic/isosteric mimic of O_2.[11] The advantage of

incorporating Xe into a macromolecular crystal is the anomalous signal of Xe. When X-ray data are collected at the absorption edge of Xe (1.5000 Å) it is possible to observe the

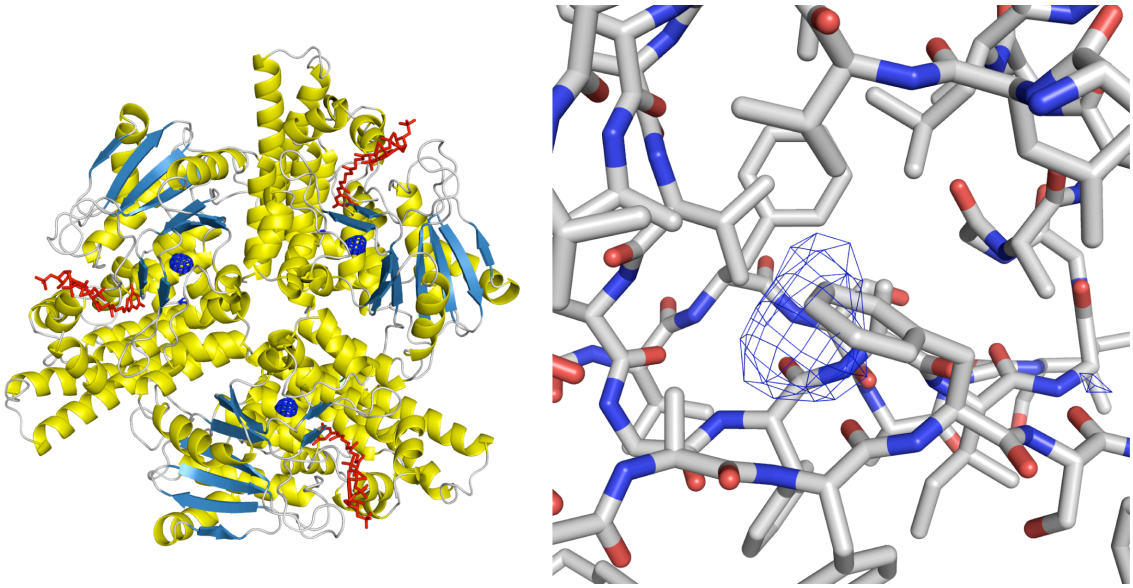


Figure 8: On the left, Xe bound to DpgC in the same location in each monomer. On the right, The hydrophobic region of DpgC where Xe binds, displacing Phe315.

noble gas bound to the protein. We soaked crystals of DpgC bound to the substrate mimic by placing them in a chamber under 300-500 psi Xe gas. We had hoped Xe would displace the oxygen bound to the hydrophobic pocket. This would be observed through an electron density map anomalous data, which would take the difference of the anomalous data from the native data, leaving only electron density for bound Xe. The experiment worked in the sense that Xe bound to the enzyme and it was observe through the difference map. However, we did not observe Xe bound to the assigned hydrophobic oxygen-binding pocket of DpgC. Instead, Xe was observed in a hydrophobic region in which the noble gas displaced a Phe residue from its original conformation. The inability of Xe to bind to the oxygen-binding pocket indicated to us that the pocket adjacent to the active site is specific for O_2 and will not promiscuously bind other hydrophobic gases.

Crystal structure of the DpgC Val425Thr mutant

To further observe the effect of our hydrophobic pocket mutants, we solved a crystal structure of one of the mutants. The V425T mutant bound to the substrate mimic

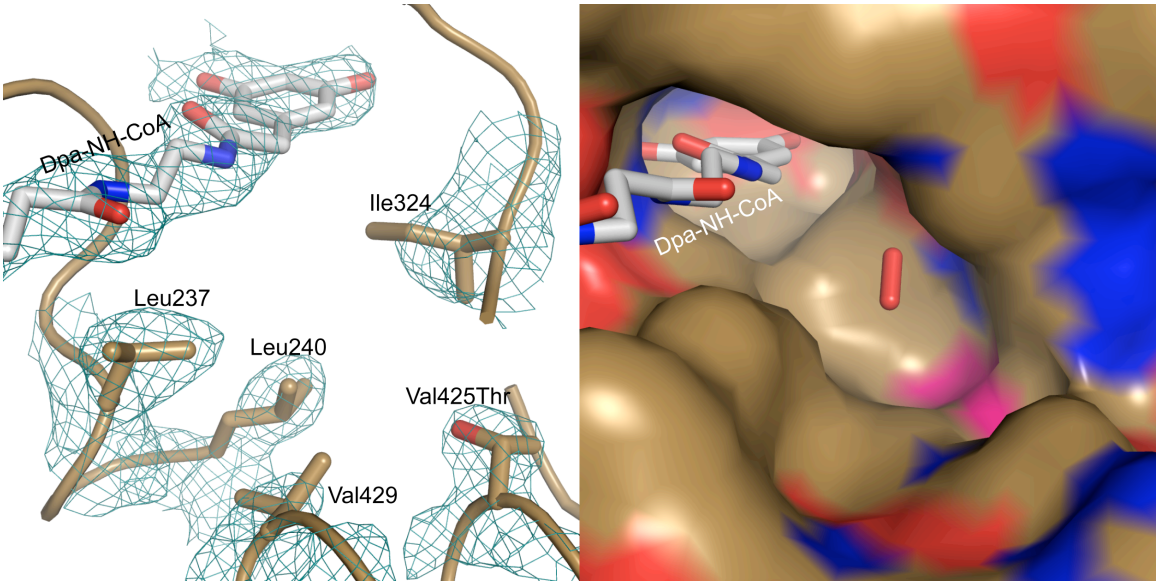


Figure 9: Electron density from composite omit map of Val425Thr mutant, on the left. On the right, surface representation of Val425Thr DpgC active site, brown represents hydrophobic region, pink represents the area that the threonine mutant disrupts the hydrophobic region. Red molecule shows oxygen position in native structure. RMSD between native and V425T = 0.6814 Å

crystallized under similar conditions to that of the native enzyme bound to the mimic. The structure of the mutant was solved through molecular replacement using the structure of the wild-type enzyme bound to the substrate mimic (Chapter 3). The overall structure of the V425T mutant is highly similar to the wild-type structure discussed in chapter 3. The same anomaly preventing the construction of the second trimer is also present in these crystals, however the quality of the electron density maps is high. Electron density surrounding the substrate mimic is clear and well defined. Dpa-NH-CoA is bound in a similar fashion to the wild-type. The key difference in the structure of the V425T mutant

is the lack of electron density in the position of the predicted bound molecular oxygen. All three trimers show no clear electron density in the location of molecular oxygen observed in the native structure. This corroborates the assignment of the orphan electron density in the native structure as molecular oxygen. Val425 is significantly far from the substrate-binding portion of the enzyme (5.89 Å). This mutant represents a subtle perturbation of the hydrophobic pocket, which based on the crystallographic data from V425T, is significant enough to prevent binding of molecular oxygen to DpgC crystals.

Synthesis and crystallization of a Dpa-fluoro-NH-CoA inhibitor

Once the starting complex of DpgC and the substrate mimic Dpa-NH-CoA was established, we designed a new mechanistic inhibitor. It was hoped that this new inhibitor would provide evidence for the next step in the reaction mechanism of DpgC. It was established through biochemical work on DpgC that the substrate goes through an enolate intermediate.[3] This intermediate is common to the crotonase superfamily,[12] so we hoped to observe an enolate intermediate bound to DpgC in a crystal structure. Since the hydrogens at the α–C position of Dpa-NH-CoA are too basic to be deprotonated by DpgC, we set about trying to lower the pKa at that position. Our strategy was to introduce a fluorine at the α–C position. We predicted the electronegative fluorine atom would lower the pKa of the remaining proton to allow deprotonation and observation of a bound enolate intermediate.

Synthesis of Dpa-α–fluoro-NH-CoA involved synthesizing amino-CoA as described in chapter 3. The fluorine is installed through chemistry outlined in Figure 10. After protection of methyl-3,5-dihydroxyphenylacetate, a regioselective bromination at the α–C position is carried out with N-bromosuccinimide. Bromine is then displaced to give a hydroxyl at the α-C position in a two-step protocol, which is then fluorinated with the DAST reagent. After deprotection, the acid is coupled to amino-CoA with PyBOP.

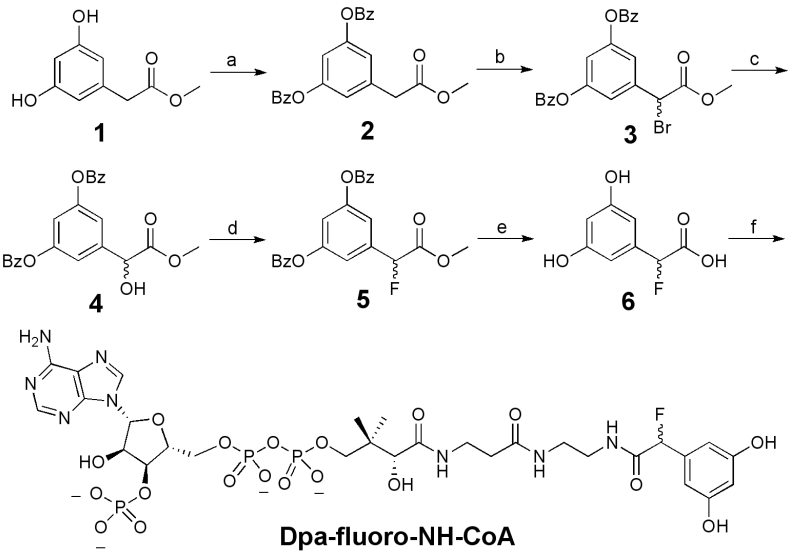


Figure 10: a. Bz-Cl, Et_3N, DCM. b. NBS, CCl_4. c. (i) NaOAc, 2:1 NMP:H_2O. (ii) HCl, MeOH. d. DAST, DCM. e. KOH, MeOH. f. PyBOP, DIPEA, 4:1 NMP:H_2O.

With the newly synthesized inhibitor in hand, crystallographic and biochemical characterization were carried out. Dpa-fluoro-NH-CoA proved to be an inhibitor with comparable kinetic parameters to Dpa-NH-CoA (Figure 10). However, assays with DpgC revealed that Dpa-fluoro-NH-CoA did not form an enolate intermediate. An NMR assay carried out in D_2O was expected to show the disappearance of the proton geminal to the fluorine (Figure 11). This result was not observed. In spite of its inability to form an

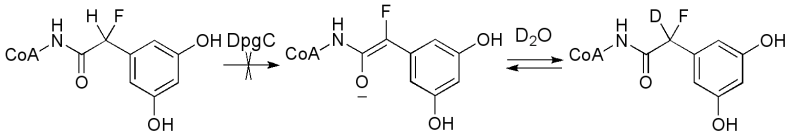


Figure 11: Assay used to biochemically determine formation of enolate by DpgC and Dpa-fluoro-AmCoA

enolate, the protein crystal structure of DpgC in complex with Dpa-fluoro-NH-CoA provided an interesting result.

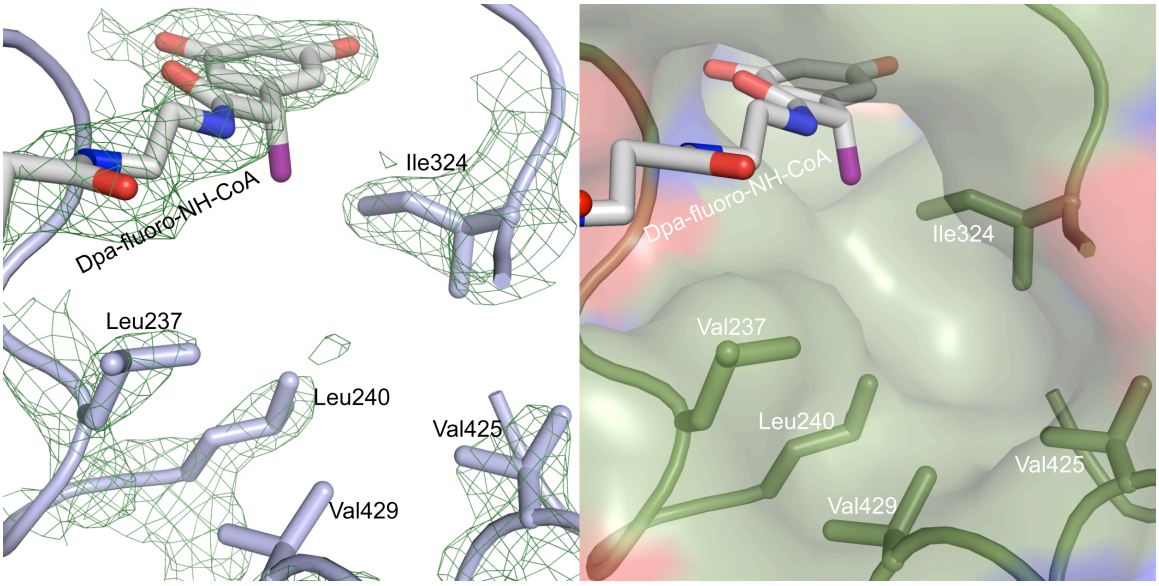


Figure 12: On the left composite omit electron density in the DpgC active site/oxygen-binding pocket. On the right, A surface representation of the DpgC oxygen-binding pocket, with hydrophobic residues in green.

The electron density for molecular oxygen was not observed in this new complex. DpgC and Dpa-fluoro-NH-CoA crystallized under similar conditions and in the same space group as DpgC bound to Dpa-NH-CoA. The structures of DpgC bound to each inhibitor are largely the same (RMSD), and electron density in the active site shows similar conformations for all residues. The major difference is the absence of molecular oxygen bound to the protein. Another notable difference is the absence of the water network anchored by Gln299.

The introduction of an electronegative fluorine atom perturbed the active site enough to prevent the binding of molecular oxygen to an otherwise identical system. The absence of molecular oxygen bound to the active shows the sensitivity of the binding pocket to minor changes.

An artificial peroxyanion hole in the structure of DpgC bound to Dpa-NH-CoA

As discussed in the introduction of this chapter, a structural motif called a peroxyanion hole has been implicated in binding molecular oxygen by proteins.[6] A peroxyanion hole functions in a similar fashion to an oxyanion hole. Protons from amides or alcohols act as hydrogen bond donors to an electron rich species. Arginine and asparagine residues have also been suggested as components of peroxyanion holes. In the case of oxygenases, molecular oxygen is the electronegative species that accepts hydrogen bonds from the enzyme and/or cofactor. While the molecular oxygen electron density in DpgC is found in a hydrophobic region of the enzyme, there are two hydrophilic components adjacent to the O_2 molecule. The backbone amide of Ile324 makes a hydrogen bond, 3.07 Å, to the bound molecular oxygen. The amide in the substrate mimic is also within hydrogen bonding distance, 3.10 Å, to molecular oxygen (Figure 13). These hydrogen bonds are oriented on either side of the observed electron density. Therefore it is reasonable to call this structural motif a peroxyanion hole.

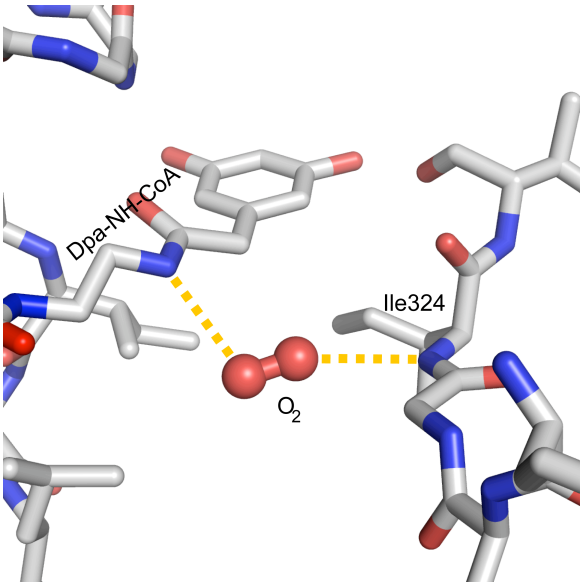


Figure 13: Peroxyanion hole formed by the substrate mimic Dpa-NH-CoA and backbone amide of Ile324.

The importance of this motif in terms of our crystallographic work is evident when considering DpgC's molecular oxygen binding constant ($K_M(O_2)$). The ambient

concentration of molecular oxygen in aqueous solution is 0.25 mM, while the $K_M(O_2)$ of DpgC is 1.7 mM. The fact that DpgC has a higher K_M than the concentration of O_2 makes the presence of molecular oxygen in the crystal structure unlikely without the help of the peroxyanion hole formed by Ile324 and the substrate mimic. From this fortuitous interaction, we were able to learn a great deal about the structural elements of DpgC that allow it to perform this fascinating chemistry.

The path of molecular oxygen in DpgC: Can we elucidate superoxide and carbon-oxygen bond formation by DpgC?

Our biochemical and crystallographic evidence support the claim that native DpgC bound to Dpa-NH-CoA shows electron density consistent with molecular oxygen. This evidence also supports the identification of a hydrophobic pocket responsible for the binding of molecular oxygen and its preparation for downstream chemistry. However, several issues remain unanswered in regard to the oxygen-binding pocket and more data needs to be collected regarding activation of molecular oxygen to superoxide. While our data indicate that the hydrophobic pocket discussed above is crucial to oxygen binding, some aspects of the DpgC crystal structure point toward other hydrophobic regions as being involved in oxygen binding and activation (Figure 14).

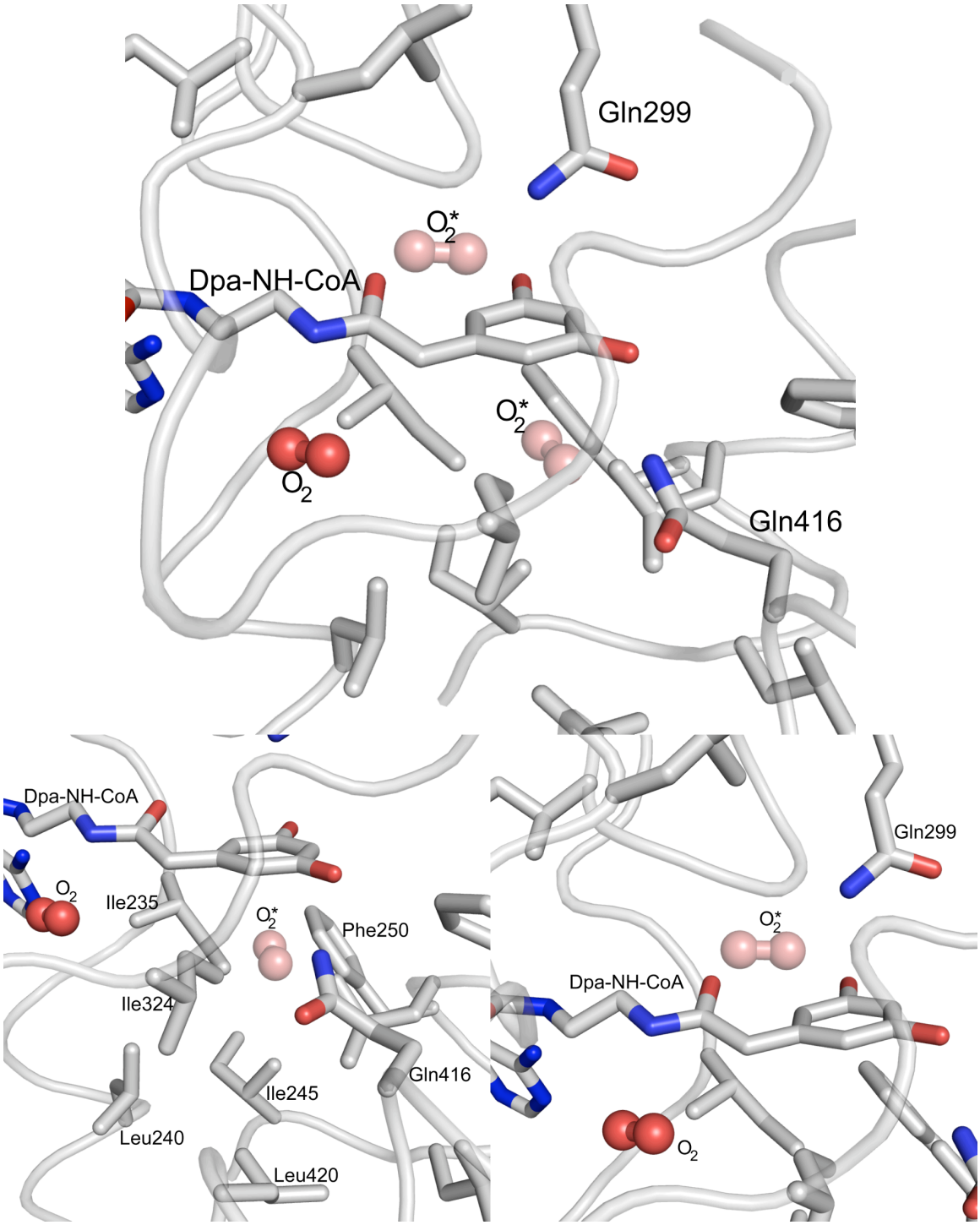


Figure 14: Possible paths of molecular oxygen in the DpgC active site. The solid red oxygen molecule is observed in the native structure bound to Dpa-NH-CoA. O_2* indicates possible positions of molecular oxygen prior to reduction by the substrate. Area below, shown bottom left, and above, shown bottom right, offer environments suitable for molecular oxygen binding and stabilization of superoxide and a peroxy-substrate intermediate (Figure 15).

Crystal structures of flavin dependent oxygenases and cofactor-independent oxygenases present structure features that are believed to stabilize the formation of superoxide and organo-peroxy intermediates.[13,14] The conformation of molecular oxygen and the substrate mimic shown in the native crystal structure could represent the complex

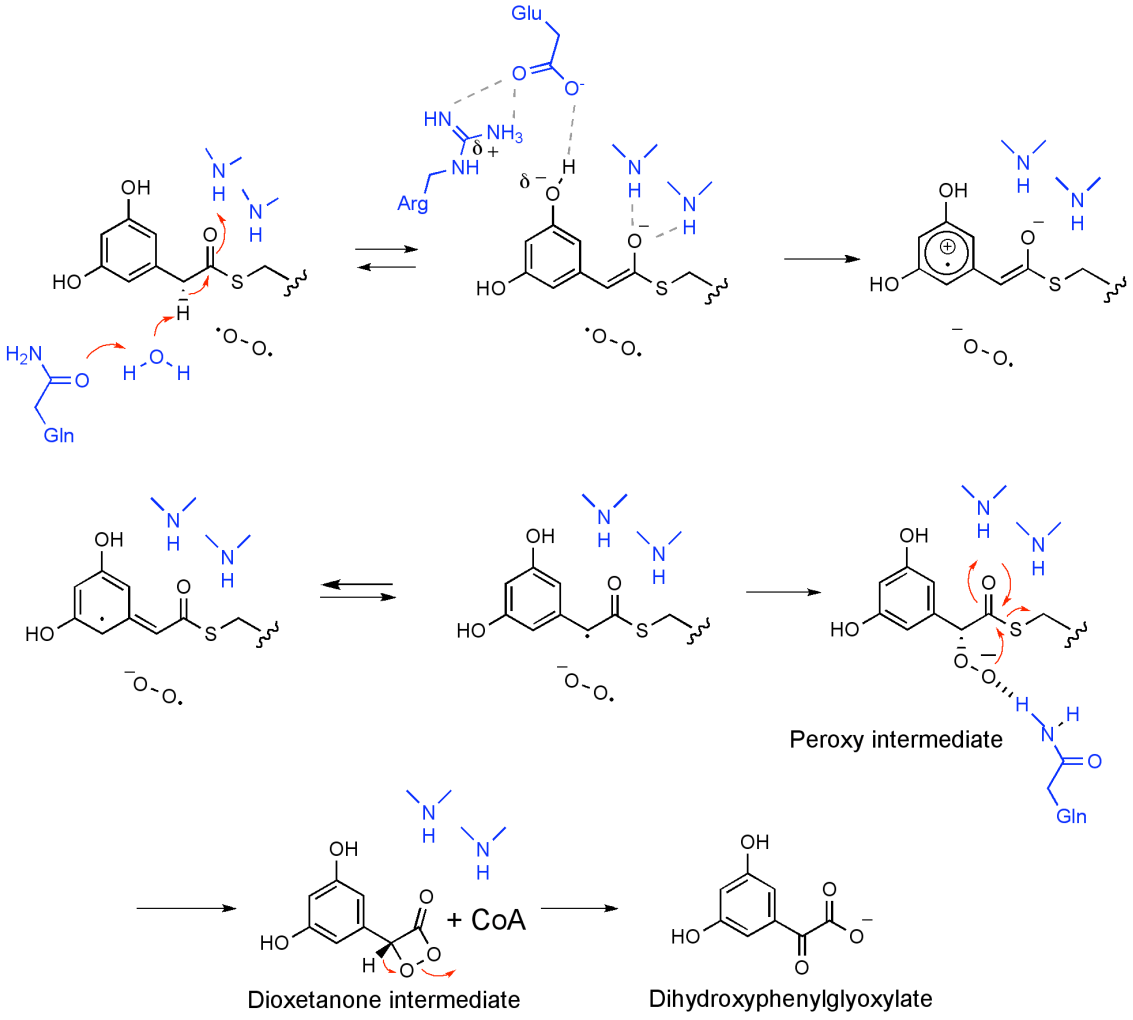


Figure 15: Proposed mechanism for DpgC: highlighted is a peroxy intermediate stabilized by a glutamine residue.

immediately before superoxide reduction. The hydrogen bond formed between molecular oxygen and the backbone amide of Ile324 presents a possible stabilizing interaction for superoxide and a possible peroxy-substrate intermediate. There are two

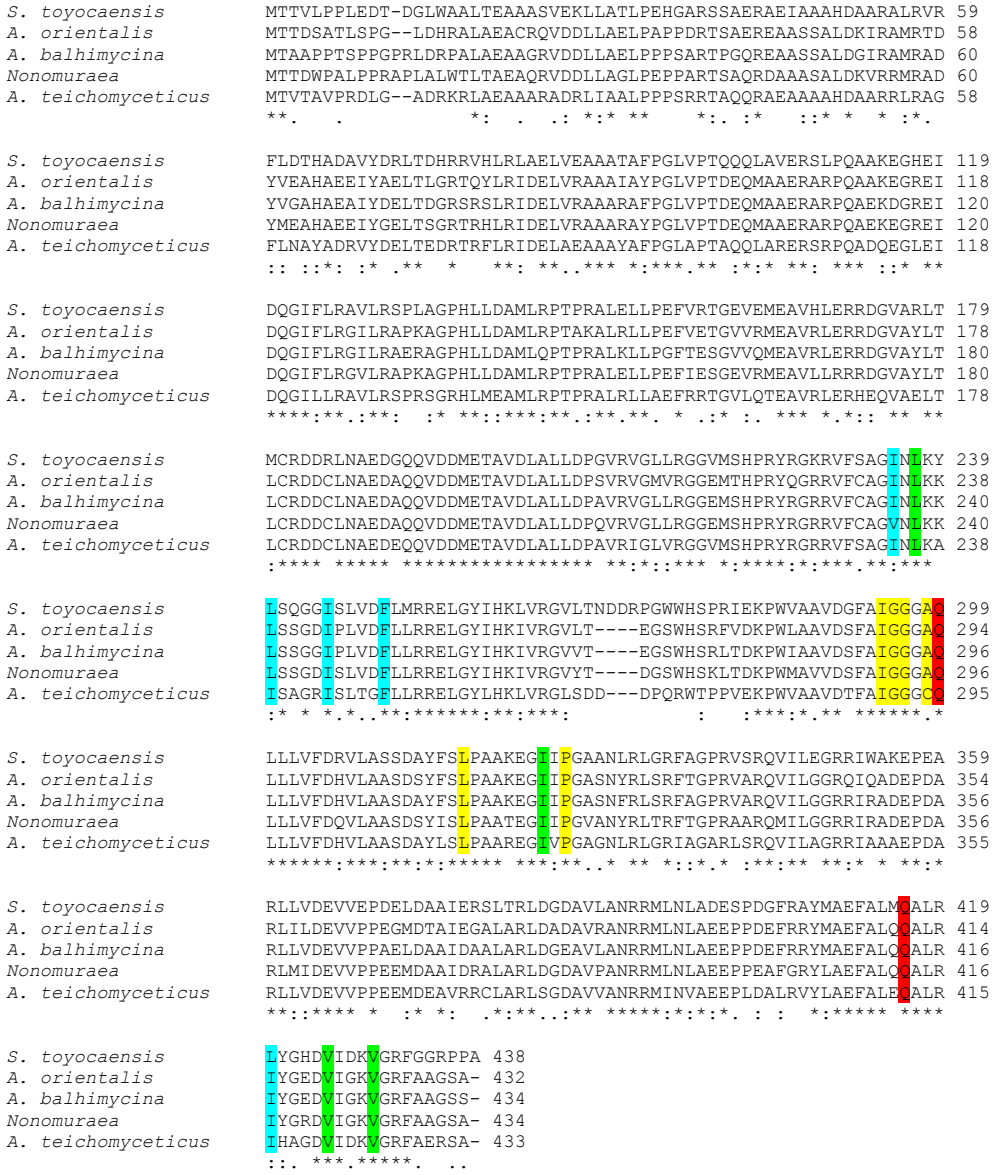


Figure 16: Alignment of five DpgC sequences from vancomycin family biosynthetic pathways. The first listed is the structure solved by Bruner lab. Highlighted in green are residues identified as a hydrophobic binding pocket from DpgC bound to Dpa-NH-CoA (Figure 5). Highlighted in cyan are residues in a hydrophobic pocket on the *si* face of the dihydroxyphenyl ring of the substrate. Highlighted in yellow are residues in a hydrophobic pocket on the *re* face of the dihydroxyphenyl ring of the substrate. Highlighted in red are glutamine residues found in the *re* and *si* hydrophobic pockets.

other possible locations for molecular oxygen to migrate to before being reduced by the substrate (Figure 14). These locations include hydrophobic regions on either side of the substrate phenyl ring. The pocket on the *si* face of the ring is lined with residues Ile235, Leu240, Ile235, Phe250, Ile324 and Phe250. This region is almost entirely made up of hydrophobic residues except for Gln416. If molecular oxygen were to bind in this pocket, it would be ideally situated to undergo single electron reduction from the electron rich ring of the substrate. The presence of Gln416 could offer some stabilization of a peroxo-intermediate (Figure 16). The *re* face of the substrate phenyl ring is adjacent to Gln299, which anchors the water network responsible for deprotonation of the C2 position on the substrate. The residues surrounding Gln299 are hydrophobic, which are hospitable for molecular oxygen migration. If oxygen were reduced to superoxide at this location, Gln299 would serve as a stabilizing residue for superoxide and the peroxy intermediate. The geometry between the substrate and Gln299 is well suited for this type of interaction. Further research is required to test the possibility of oxygen migration from its current position in the native bound structure to a position on either side of the substrate phenyl ring. The residues in the hydrophobic regions on either side of the substrate phenyl ring are conserved in all DpgC homologues.

Establishing how DpgC facilitates a single-electron transfer from the substrate to molecular oxygen and how a peroxy-intermediate and a possible dioxetanone intermediate are formed and stabilized are key questions that must be answered to fully understand the mechanism of this enzyme.

Materials and Methods

Mutagenesis. Point mutations of DpgC were made using the QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene: La Jolla, CA). Elisha Fielding carried out the hydrophobic pocket mutagenesis work. The procedure is detailed in published material.[2]

DpgC substrate kinetic parameters. Elisha Fielding carried out substrate kinetic parameter measurements. Kinetic parameters of the DpgC active site mutants were analyzed using the DTNB reporter assay as described above with one modification: the concentration of the DpgC mutant was 0.186 μM.

Measurement of kinetic parameters for dioxygen. All enzymes were prepared for assays using the procedure described above. O_2 consumption was measured using a Clark-type O_2 electrode (Hansatech Instruments) and the electrode signal was recorded using Virtual Bench Data Logger. The electrode was calibrated using the 2,3-dihydroxybiphenyl and DHBD.[15] Data was collected every 0.003 s and initial velocities were determined from progress curves (Microsoft Excel). Steady-state rate equations were fir to data using the least squares and dynamic weighting options of LEONORA.[16] The kinetic parameters of DpgC with respect to O_2 were determined in 100 μM Dpa-CoA and varying concentrations of O_2. Reaction buffers (20 mM TrisHCl, 50 mM NaCl, pH 7.5) at 25^{o}C were prepared by bubbling mixtures of O_2 and N_2 gases for at least 5 min. The gases were mixed using a gas proportioner (Specialty Gas Equipment) and transferred to the reaction vessel using a Hamilton syringe. The concentration of O_2 was confirmed using the O_2 electrode. The assay was initiated by injection of enzyme (2 μM) into the reaction vessel containing buffer and Dpa-CoA (1 mL final volume). The initial rates increased almost linearly with O_2 concentration and no points above the K_m could be obtained, causing the low precision in the kinetic parameters.

Xenon Crystal Derivatives. DpcC-inhibitor co-complex crystals were placed in a Xenon pressure chamber (Hampton Research) for a range of 5 to 45 min with pressures of 300 to 500 psi. The xenon equilibrated crystals were flash frozen within seconds of removal from the xenon chamber.

Enzyme Production. Overexpression and purification of DpgC and all DpgC mutants were carried out under conditions identical to those described in chapter 2.

Crystallization and data collection. Recombinant DpgC (48kDa, 438 amino acids) from *Streptomyces toyocaensis* was crystallized in the presence of Dpa-NH-CoA or Dpa-α–fluoro-NH-CoA by the hanging drop vapor diffusion method at 20^{o}C. 120 mL DpgC (12 mg/mL in 20mM Tris, 100mM NaCl, 1mM bME, pH 7.5) was mixed with 30 mL 8.73mM Dpa-NH-CoA or Dpa-fluoro-NH-CoA and incubated for 30 min. at 4^{o}C. 1 mL of this solution was added to 1 mL of reservoir solution (100mM sodium citrate, 150mM ammonium acetate and 12-20% PEG 4,000, pH 5.6). Prism like crystals appeared after two days. Crystals were transferred to a cryoprotectant solution of reservoir solution with 20% glycerol, briefly soaked then flash frozen in liquid nitrogen. X-ray diffraction data for DpgC bound to Dpa-fluoro-NH-CoA was collected on the X29A beamline at the National Synchrotron Light Source at Brookhaven National Labs on ADSC Q315r Crystal Logic diffractometer. Data for DpgC Val425Thr bound to Dpa-NH-CoA was collected on the X12C beamline at the National Synchrotron Light Source at Brookhaven National Labs on an ADSC Q210 Crystal Logic diffractometer. Data was collected at 100^{o} K. Diffraction intensities were indexed, integrated and scaled with HKL2000 as

summarized in Table 1. The crystal belongs to the space group $P2_12_12$ with unit cell dimensions: a=39.155, b=156.079, c=170.959; α, β,γ=90°.

Model building, refinement and graphics. The structures were obtained by simple molecular replacement with the inhibitor bound structure of DpgC (PDB code: 2NP9). The model was built using the program COOT.[17] Refinement cycles and generation of $2F_o$-F_c and F_o-F_c composite omit maps were done using the CNS suite of programs.[18] Cycles of rigid body refinement, simulated annealing, composite omit maps and addition of waters into the structure was done until R_{free} values were no longer improving to any extent. To generate images of the structure the program Pymol was used.

Table 1. Crystallography data collection and refinement statistics

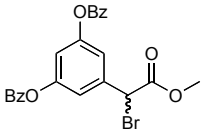
Data Collection	Dpa-α–fluoro-NH-CoA	Val425Thr
Space Group	$P2_12_12$	$P2_12_12$
Cell dimensions (Å)		
a, b, c (Å)	139.952, 156.079, 170.959	139.155,156.141,171.284
α, β, γ (°)	90, 90, 120	90, 90, 120
Wavelength (Å)	1.00000	1.00000
Resolution (Å)	2.5	2.8
R_{merge} or R_{sym}	0.081 (0.324)	0.138 (0.544)
$I/\sigma I$	22.9 (3.4)	18.2 (4.1)
Completeness	99.2 (96.4)	98.9 (98.8)

Redundancy	6.1 (5.2)	7.4 (7.1)
Refinement		
Resolution (Å)	2.5	2.8
No. reflections	128431	91225
R_{work}/R_{free}	0.336/0.362	0.339/0.344
No. atoms	10154	9846
B-factors	38.2	44.5
R.M.S.D.		
Bond lengths (Å)	0.029	0.032
Bond angles (^{o})	1.9	2.1

Synthesis of Dpa-fluoro-amino-CoA

^{1}H NMR spectra were recorded using a Varian Unity INOVA 400 MHz (400 MHz) spectrometer. Chemical shifts are reported in ppm from trimethylsilane with the solvent as the internal standard ($CDCl_3$: δ 7.26 ppm, D_2O: δ 4.80 ppm). Data are reported as follows: chemical shift, (multiplicity [singlet (s), doublet (d), triplet (t), quartet (q) and multiplet (m)], coupling constants [Hz], and integration. ^{13}C NMR spectra were recorded on a Varian Unity INOVA 400 MHz (100 MHz) spectrometer with complete proton decoupling. Chemical shifts are reported with the solvent as the internal standard ($CDCl_3$: δ 77.23 ppm). High resolution mass spectrometry (MS) was performed at the Mass Spectrometry Facility at Boston College on a LCT ESI-MS.

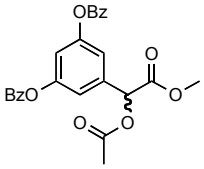
Chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO). Deuterated NMR solvents were purchased from Cambridge Isotope (Andover, MA). Purification of intermediates was performed on silica gel 60 from VWR (West Chester, PA). All work-up and purification procedures were performed in air using solvents purchased from Fisher Scientific (Pittsburgh, PA).



Methyl-2-bromo-2-(3,5-dibenzoylphenyl)acetate(3): To a solution of methyl 3,5-dihydroxyphenylacetate (0.519g, 2.84 mmol) in CH_2Cl_2 (20 mL) was added diisopropylethylamine (2.1mL, 11.4 mmol) and benzoyl chloride (1.7 mL, 11.4 mmol). The mixture was stirred at room temperature for 16h before washing twice with saturated sodium bicarbonate solution (2x10mL). The organic layer was separated and dried *in vacuo*.

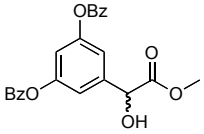
The above crude product was dissolved in 15 mL dry CCl_4 followed by addition of *N*-bromo-succinimide(1.2g, 6.72 mmol). The mixture was refluxed at 80^{o}C for 4 hours. The reaction was washed with water (10mL) followed by brine washes (2x10mL). The reaction was purified on silica in 7:3 hexanes: ethyl acetate to give the desired product as a white powder(1.113g, 83.5% for 2 steps).

^{1}H NMR ($CDCl_3$, 400 MHz) δ 8.195 (d, J = 9.6Hz, 4H), 7.657 (t, J = 16.4Hz 2H), 7.521 (t, J = 16 Hz, 4H), 7.415 (s, 2H), 7.410 (s, 1H), 5.375 (s, 1H), 3.816 (s, 3H). **^{13}C NMR** ($CDCl_3$ 100Hz) δ 168.604, 164.849, 151.804, 138.261, 134.318, 130.647, 129.089, 120.034, 117.293, 54.055, 45.413. **HRMS** (ESI+) Calcd for $C_{23}H_{17}O_6NaBr$ 491.0106, found 491.0103.



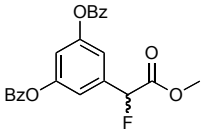
Methyl-2-acetyl-2-(3,5-dibenzoylphenyl)acetate(4): To a solution of **3** (0.737g, 1.576 mmol) in 2:1 NMP:H_2O(20mL) was added sodium acetate(0.280g, 2.06mmol). The mixture was refluxed for 3 hours at 80^oC. The reaction was quenched by addition of 1:1 ethyl acetate: water(20mL). The organic layer was concentrated *in vacuo* and purified on silica in 7:3 hexanes: ethyl acetate. The desired product was obtained as a white solid(0.490g, 1.093mmol, 69.4%)

^{1}H NMR ($CDCl_3$, 400 MHz) δ 8.2 (d, J = 9.6Hz, 4H), δ 7.658 (t, J = ?, 2H), 7.526 (t, J = 16 Hz, 4H), 7.326 (s, 2H), 7.320 (s, 1H), 5.980 (s, 1H), 3.771 (s, 3H), 2.230 (s, 3H). **^{13}C NMR** ($CDCl_3$ 100Hz) δ 170.133, 168.715, 164.663, 151.650, 136.3119, 134.023, 130.368, 128.810, 118.493, 116.919, 73.670, 53.098, 20.877. **HRMS** (ESI+) Calcd for $C_{25}H_{20}O_8Na$ 471.1056, found 471.1048

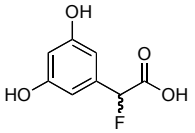


Methyl-2-hydroxy-2-(3,5-dibenzoylphenyl)acetate(5): Compound **4** (227mg, 0.481mmol) was dissolved in 12mL 5:1 methanol:CH_2Cl_2, to this solution was added 2 mL concentrated HCl. The reaction was allowed to stir at room temperature for 16h. The crude product was purified on silica in 7:3 hexanes:ethyl acetate to yield the desired product as a clear oil (105mg, 0.245mmol, 50.9%).

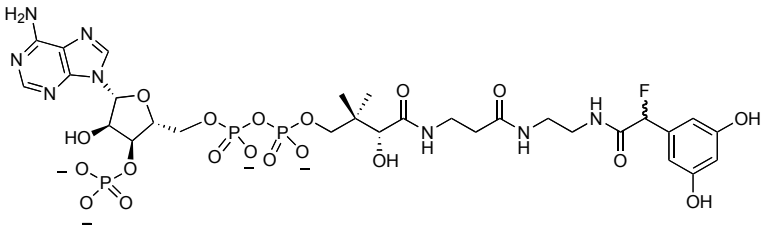
^{1}H NMR ($CDCl_3$, 400 MHz) δ 8.25 (d, J = 9.6Hz, 4H), δ 7.697 (t, J = 14.8, 2H), 7.566 (t, J = 15.6 Hz, 4H), 7.368 (s, 2H), 7.265 (s, 1H), 5.321 (d, J = 5.2, 1H), 3.793 (s, 3H). **^{13}C NMR** ($CDCl_3$ 100Hz) δ 173.28, 164.73, 151.52, 140.86, 133.91, 130.31, 129.26, 128.75, 117.48, 115.93, 72.20, 53.48. **HRMS** (ESI+) Calcd for $C_{23}H_{18}O_7Na$ 429.0950, found 429.0944



Methyl-2-fluoro-2-(3,5-dibenzoylphenyl)acetate(6): A solution of compound **5** (0.1816 g, 0.447 mmol) in 10 mL CH_2Cl_2 was cooled to 0°C on ice. 50 µL of DAST (0.122 g, 0.7568 mmol) was added and the reaction was allowed to warm to room temperature. After stirring for 2.5 hours, the reaction was concentrated in vacuo and purified on silica in 6:4 hexanes:ethyl acetate to give the desired product as a cloudy, white oil (0.384 g, 0.89 mmol, 85%). **^{1}H NMR** ($CDCl_3$, 400 MHz) δ 8.1955 (d, J = 10 Hz, 4H), 7.649 (t, J = 7.2 Hz, 2H), 7.526 (t, J = 16 Hz, 4H), 7.312 (s, 2H), 7.285 (s, 1H), 5.868 (d, J = 47.2 Hz, 1H), 3.827 (s, 3H). **^{13}C NMR** ($CDCl_3$ 100Hz) δ 164.669, 151.781, 136.694, 136.476, 134.076, 130.421, 128.847, 117.307, 89.410, 87.539, 53.159. **HRMS** (ESI+) Calcd for $C_{23}H_{17}O_6NaF$ 431.0907, found 431.0893



2-(3.5-dihydroxyphenyl)-2-fluoroacetic acid(7): Compound **6**(0.0426mmol, 17.4 mg) was added to a solution of KOH (0.256 mmol, 14.35 mg) in methanol. The mixture refluxed for 4 hours. Purification on reverse phase(C18) HPLC gave the product as an orange oil (6.4 mg, 81.01%). HPLC buffers: A: 0.1% TFA in H_2O; B: Acetonitrile; gradient: 0-5% B over 50 minutes. Product elutes at 35 minutes. **^{1}H NMR** (D_2O, 400MHz) δ 6.55 (s, 2H), 6.432 (s, 2H), 5.689 (d, J=48.8, 1H). **HRMS** (ESI+) Calcd for $C_8H_6O_4F$ 185.250, found 185.253

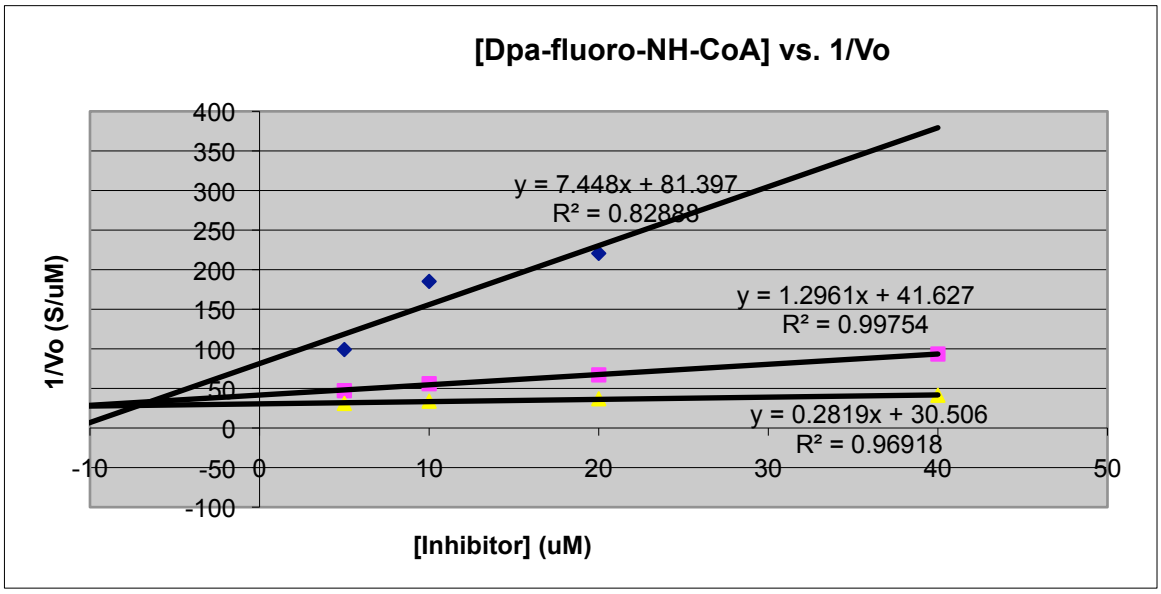


DPA-fluoro-NH-CoA: Amino-CoA(4.69mmol, 3.5mg), compound **7**(16.13

mmol, 3.0 mg) and PyBOP (19.21 mmol, 10.0 mg) were dissolved in 665 mL 4:1 DMF:H_2O. Diisopropylethylamine (82.1 mmol, 7.8775 mL) was then added. The reaction was gently shaken at room temperature for 3h. The reaction was concentrated *in vacuo* until dryness. The compound was purified on HPLC using a linear gradient of 0-30% acetonitrile over 40 min in 0.1% trifluoroacetic acid. The product eluted at 21min, and gave 3.54 μmol (75.5% yield).

^{1}H NMR (D_2O, 400 MHz) δ 8.59 (s, 1H), 8.39 (s, 1H), 6.42 (s, 2H), 6.33 (2, 1H), 6.21 (d, 1H), 5.70 (d, J = 46.8, 1H), 4.43 (br s, 1H), 4.29 (br s, 2H), 4.01 (overlapping m, 2H) 3.84 (m, 2H), 3.61 (m 2H), 3.35 (m 2H), 2.33 (t, J = 13.6, 2H), 0.936 (s, 3H), 0.828 (s, 3H). **HRMS** (ESI-) *m/z* Calcd for $C_{29}H_{41}N_8O_{19}FP_3$ 917.1685, found 917.1646.

Kinetic Inhibition Assays with Analog 4. Kinetic analysis for DpgC was done using the DTNB [5,5'-dithiobis(2-nitrobenzoic acid)] reporter assay as described[10] with one modification: the concentration of DpgC was 0.4 μM. To determine the kinetic parameters of DpgC for the substrate Dpa-CoA, a freshly prepared solution of DTNB (1 mM), Tris·HCl (250 mM), and increasing concentrations of DPA-CoA was mixed with DpgC (0.4 μM) at 24 ºC (1 mL). The reactions were monitored in a UV-Vis spectrophotometer (Agilent 8453. Agilent Technologies: Santa Clara, CA) at 412 nm for 5 or 3 min. The observed initial rate of absorbance increase was converted to initial reaction velocity V_0 (ε_{412nm}=13,600 $M^{-1}\cdot cm^{-1}$). $1/V_0$ was plotted as a function of 1/[S] in a Lineweaver and Burk plot. The measured K_M was 4.7 μM and the k_{cat} was 7.8 min^{-1} (similar to literature values: K_M= 6 μM, k_{cat}= 10 min^{-1}). [4]



Inhibition assays were performed using the same conditions as above with increasing concentrations of analog Dpa-fluoro-NH-CoA at a constant concentration of substrate Dpa-CoA ([S] = 2 μM, 10 μM and 40 μM). $1/V_0$ was plotted as a function of concentration of 11 (Figure S3) according to the method of Dixon.[19,20] The K_i was established to be 7.876 ±0.652 μM and the plots indicate competitive inhibition.

The relative enzyme activity (initial velocity with inhibitor over initial velocity without inhibitor, when [S]=5 μM) was plotted as a function of concentration of inhibitor 4. The concentration which inhibited 50% of the enzyme activity was taken as the IC_{50}. The IC_{50} of 4 is 5.3 μM when the substrate concentration is 5.0 μM.

References:

1 Fetzner, S., Oxygenases without requirement for cofactors or metal ions. *Appl Microbiol Biotechnol* **60** (3), 243 (2002).

2 Widboom, P. F., Fielding, E. N., Liu, Y., and Bruner, S. D., Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis. *Nature* **447** (7142), 342 (2007).

3 Tseng, C. C., Vaillancourt, F. H., Bruner, S. D., and Walsh, C. T., DpgC is a metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase in the vancomycin biosynthetic Pathway. *Chem Biol* **11** (9), 1195 (2004).

4 Bellelli, A., Morpurgo, L., Mondovi, B., and Agostinelli, E., The oxidation and reduction reactions of bovine serum amine oxidase. A kinetic study. *Eur J Biochem* **267** (11), 3264 (2000).

5 Klinman, J. P., Life as aerobes: are there simple rules for activation of dioxygen by enzymes? *J Biol Inorg Chem* **6** (1), 1 (2001).

6 Taga, M. E. et al., BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature* **446** (7134), 449 (2007).

7 Eswaramoorthy, S., Bonanno, J. B., Burley, S. K., and Swaminathan, S., Mechanism of action of a flavin-containing monooxygenase. *Proc Natl Acad Sci U S A* **103** (26), 9832 (2006).

8 Gatti, D. L., Entsch, B., Ballou, D. P., and Ludwig, M. L., pH-dependent structural changes in the active site of p-hydroxybenzoate hydroxylase point to the importance of proton and water movements during catalysis. *Biochemistry* **35** (2), 567 (1996).

9 Malito, E., Alfieri, A., Fraaije, M. W., and Mattevi, A., Crystal structure of a Baeyer-Villiger monooxygenase. *Proc Natl Acad Sci U S A* **101** (36), 13157 (2004).

10 Chen, H., Tseng, C. C., Hubbard, B. K., and Walsh, C. T., Glycopeptide antibiotic biosynthesis: enzymatic assembly of the dedicated amino acid monomer (S)-3,5-dihydroxyphenylglycine. *Proc Natl Acad Sci U S A* **98** (26), 14901 (2001).

11 Wentworth, P., Jr. et al., Antibody catalysis of the oxidation of water. *Science* **293** (5536), 1806 (2001).

12 Holden, H. M., Benning, M. M., Haller, T., and Gerlt, J. A., The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme a thioesters. *Acc Chem Res* **34** (2), 145 (2001).

13 Sciara, G. et al., The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis. *Embo J* **22** (2), 205 (2003).

14 Roth, J. P. and Klinman, J. P., Catalysis of electron transfer during activation of O2 by the flavoprotein glucose oxidase. *Proc Natl Acad Sci U S A* **100** (1), 62 (2003).

15 Vaillancourt, F. H. et al., Molecular basis for the stabilization and inhibition of 2, 3-dihydroxybiphenyl 1,2-dioxygenase by t-butanol. *J Biol Chem* **273** (52), 34887 (1998).

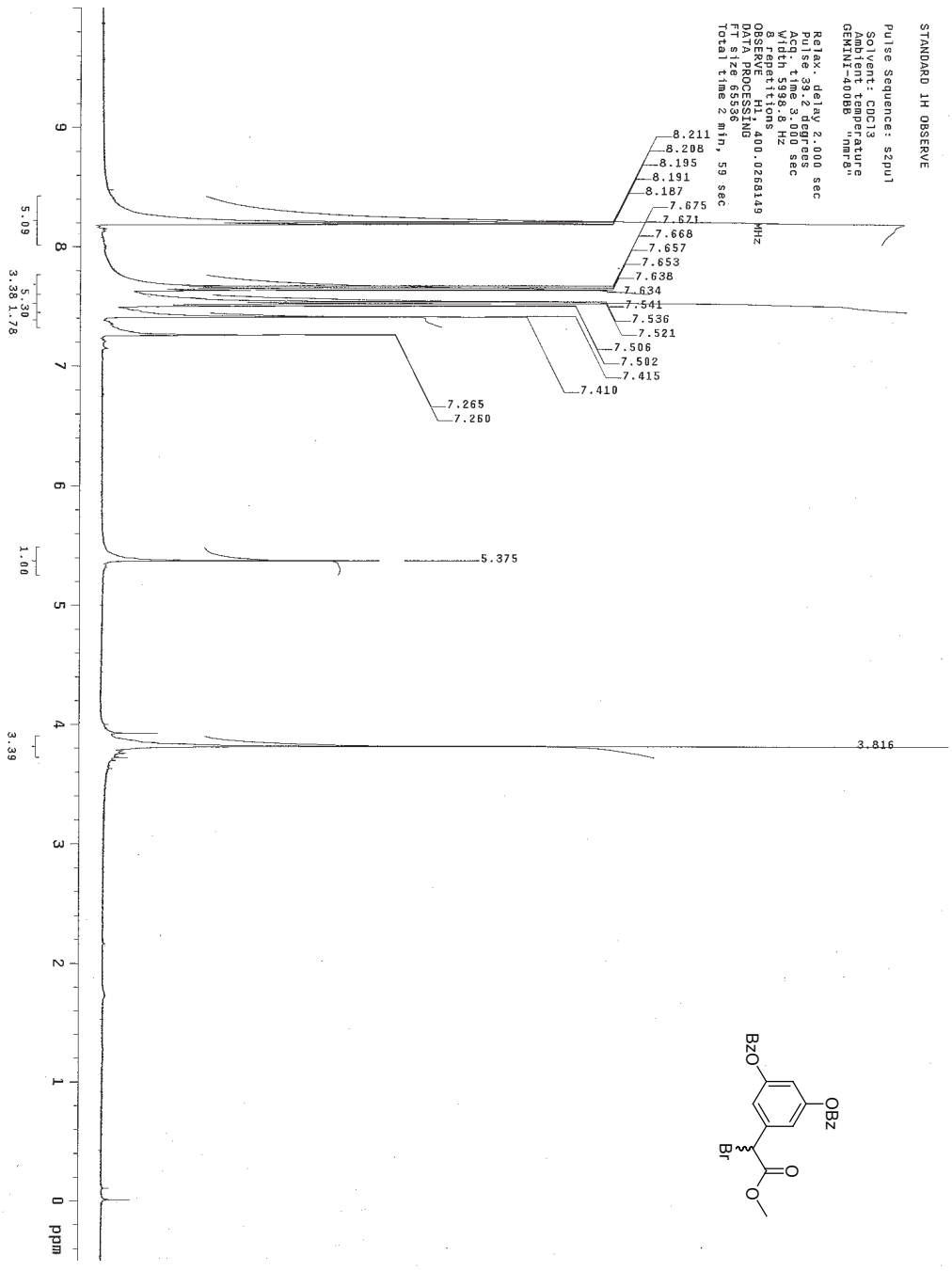
16 Cornish-Bowden, A., *Analysis of Enzyme Kinetic Data*. (Oxford University Press, New York, 1995).

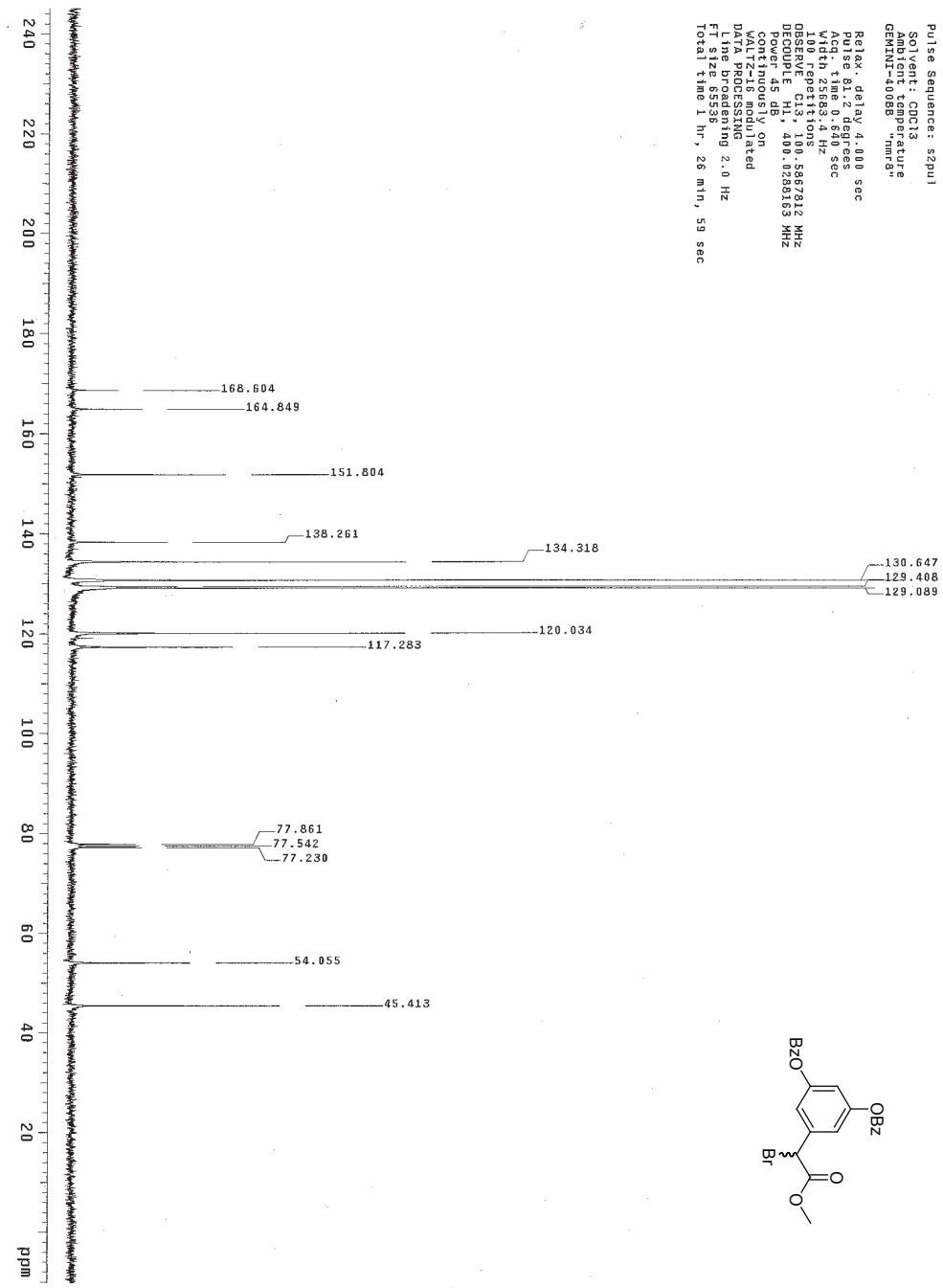
17 Emsley, P. and Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60** (Pt 12 Pt 1), 2126 (2004).

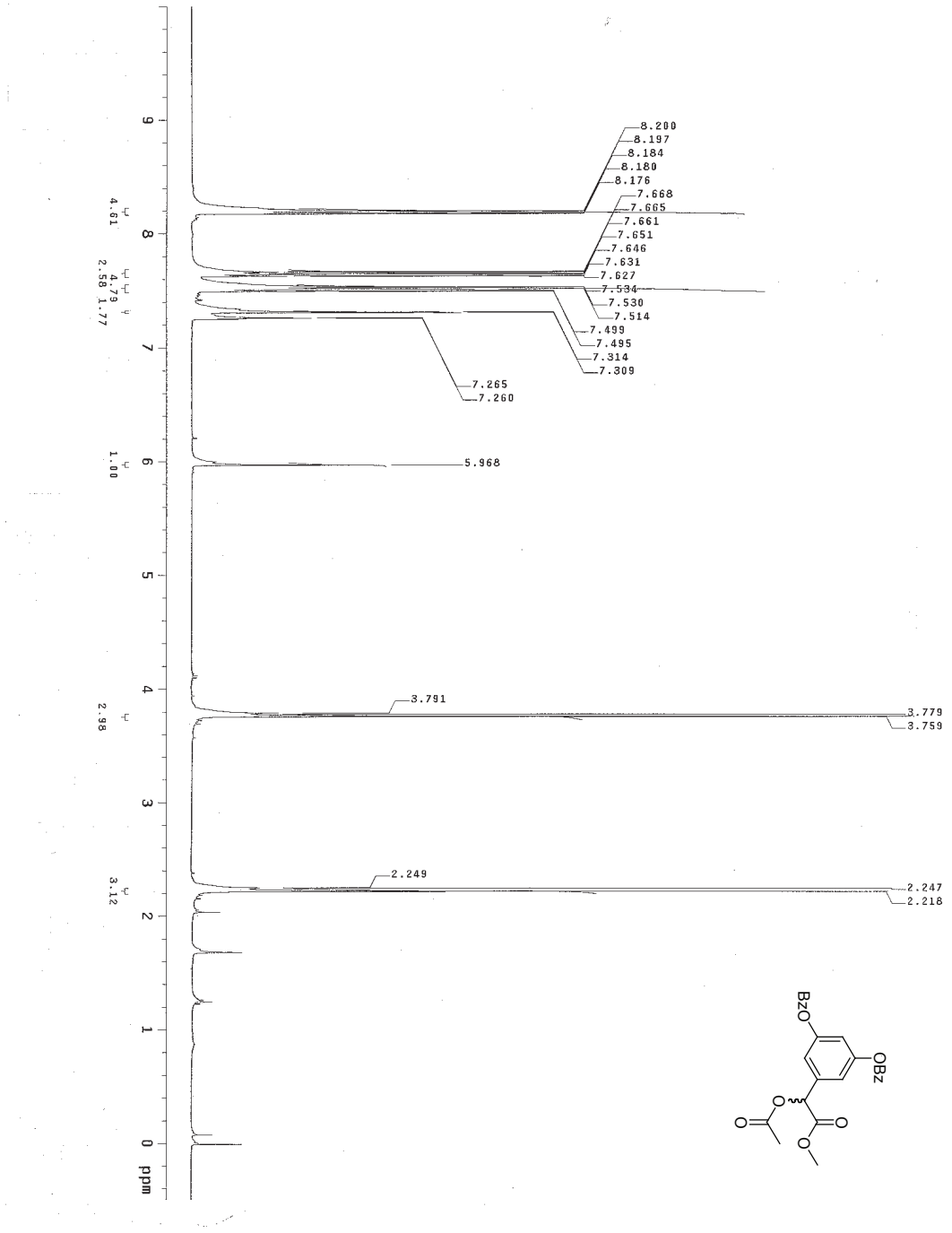
18 Brunger, A. T. et al., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54** (Pt 5), 905 (1998).

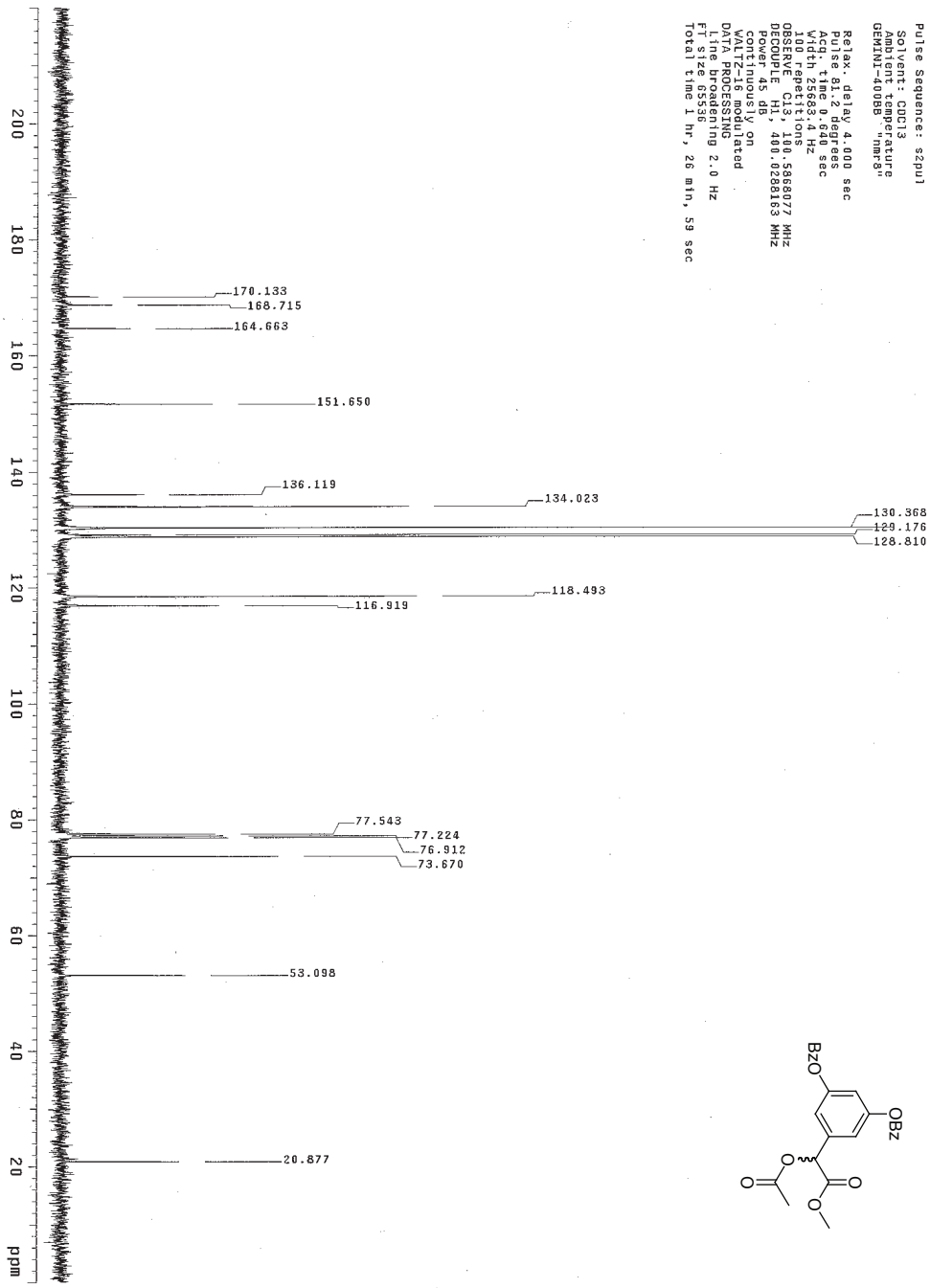
19 B.T. Burlingham, T.S. Widlanski. *Journal of Chemical Education* **80**, 214 (2003).

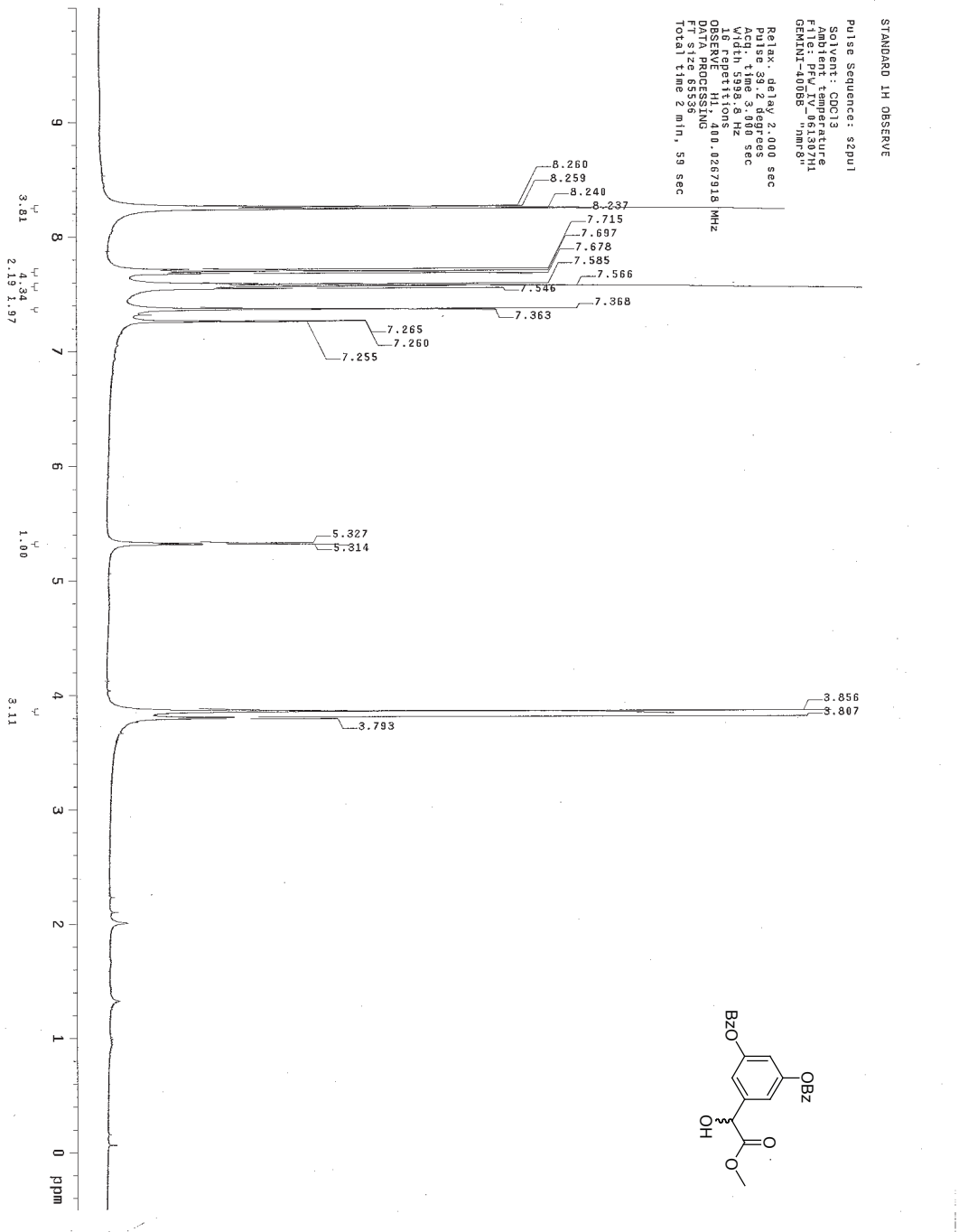
20 Dixon, M., The determination of enzyme inhibitor constants. *Biochem J* **55** (1), 170 (1953).

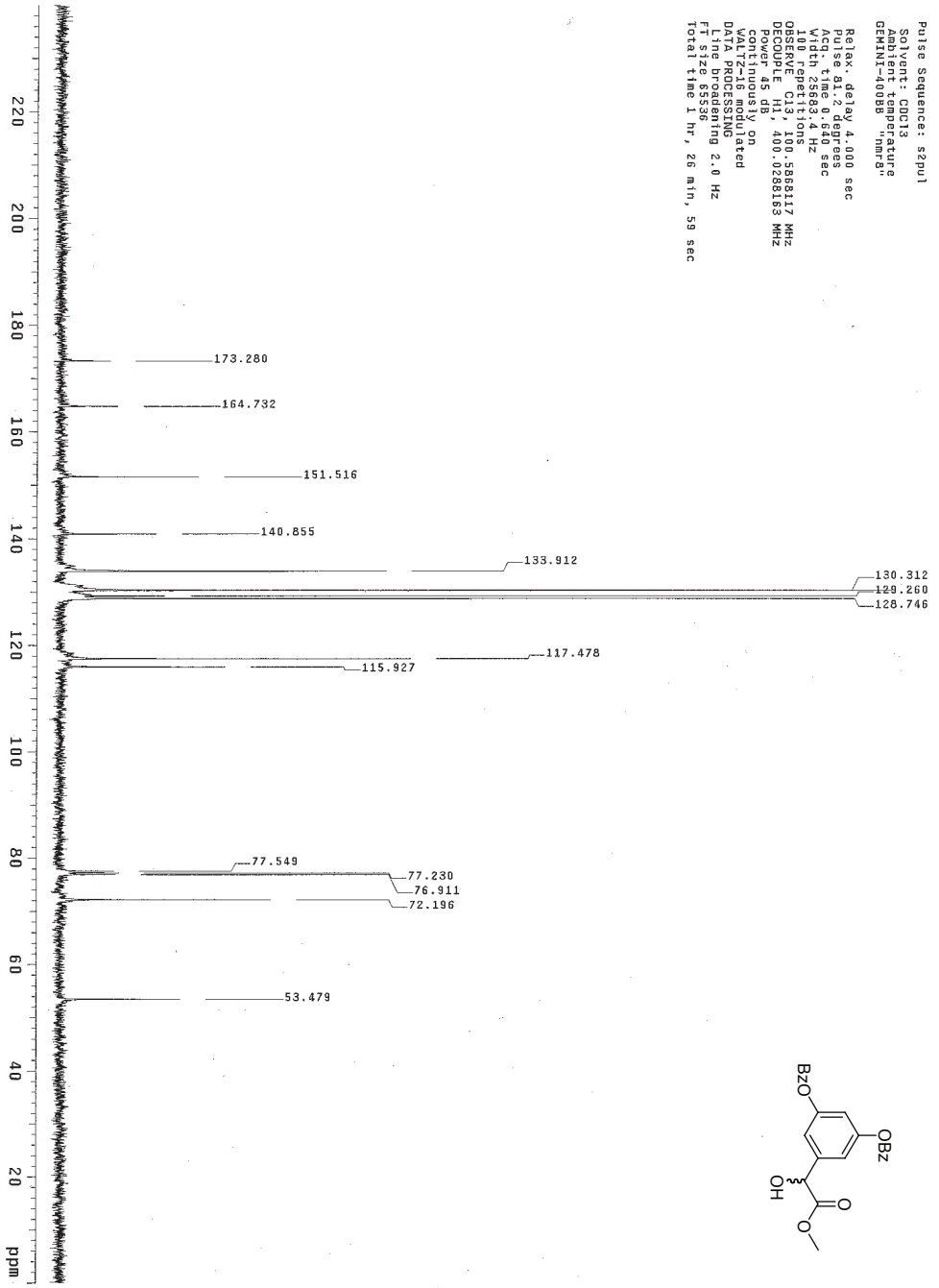

STANDARD 1H OBSERVE
Pulse Sequence: s2pul
Solvent: CDCl3
Ambient temperature
GEMINI-400BB "nmrg"
Relax. delay 2.000 sec
Pulse 39.2 degrees
Acq. time 3.000 sec
Width 5998.8 Hz
8 repetitions
OBSERVE H1, 400.0268149 MHz
DATA PROCESSING
FT size 65536
Total time 2 min, 59 sec
8.211
8.208
8.195
8.191
8.187
7.675
7.671
7.668
7.657
7.653
7.638
7.634
7.541
7.536
7.521
7.506
7.502
7.415
7.410
7.265
7.260
5.375
3.816
5.09
3.38
5.30
1.78
1.00
3.39
ppm
BzO
OBz
Br
O
O

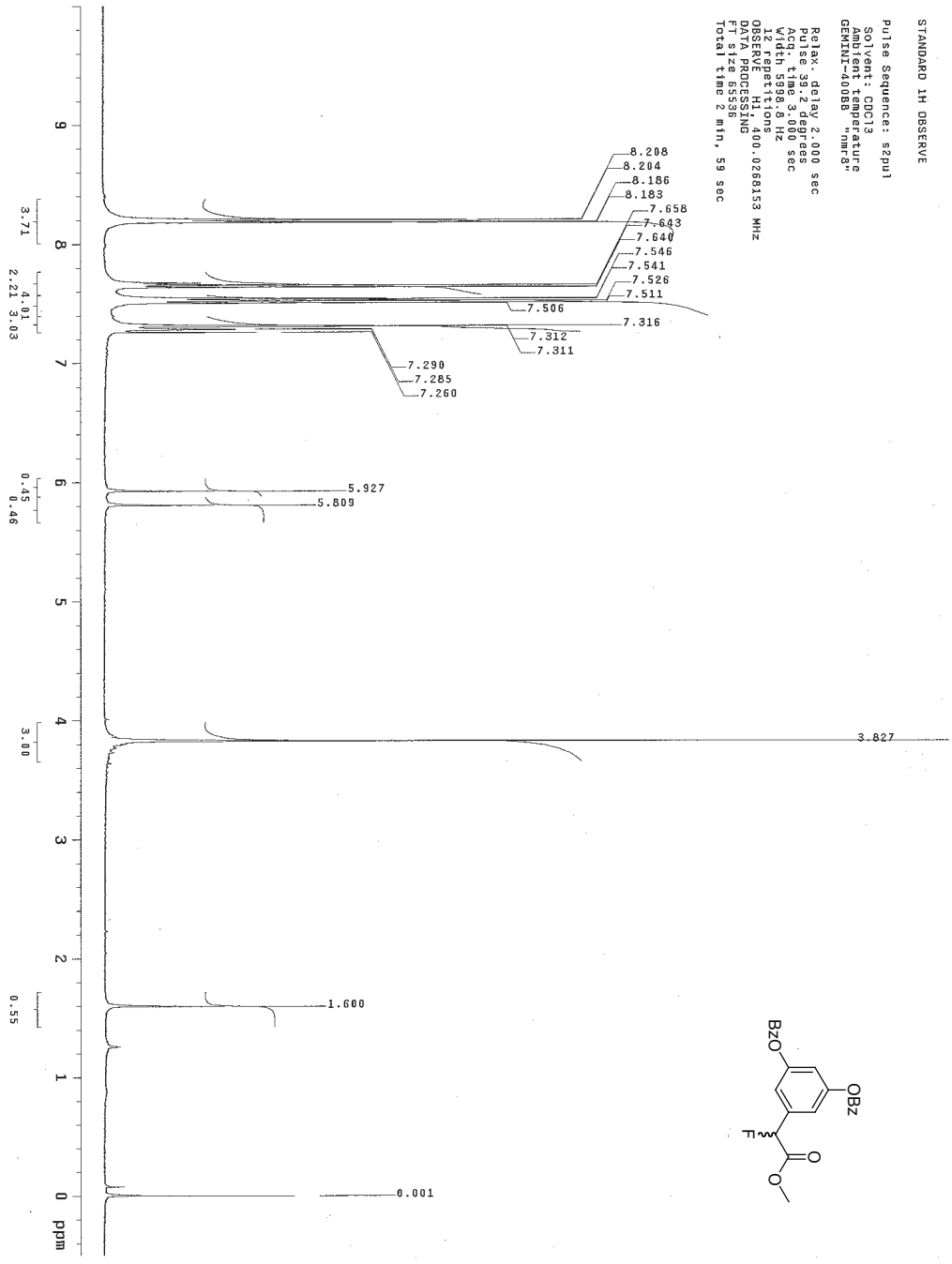

13C OBSERVE
Pulse Sequence: s2pul
Solvent: CDCl3
Ambient temperature
GEMINI-400BB "nmr8"
Relax. delay 4.000 sec
Pulse 81.2 degrees
Acq. time 0.640 sec
Width 25683.4 Hz
100 repetitions
OBSERVE C13, 100.5867812 MHz
DECOUPLE H1, 400.0288163 MHz
Power 45 dB
continuously on
WALTZ16 modulated
DATA PROCESSING
Line broadening 2.0 Hz
FT size 65536
Total time 1 hr, 26 min, 59 sec
168.604
164.849
151.804
138.261
134.318
130.647
129.408
129.089
120.034
117.283
77.861
77.542
77.230
54.055
45.413
ppm
BzO
OBz
Br
O
O

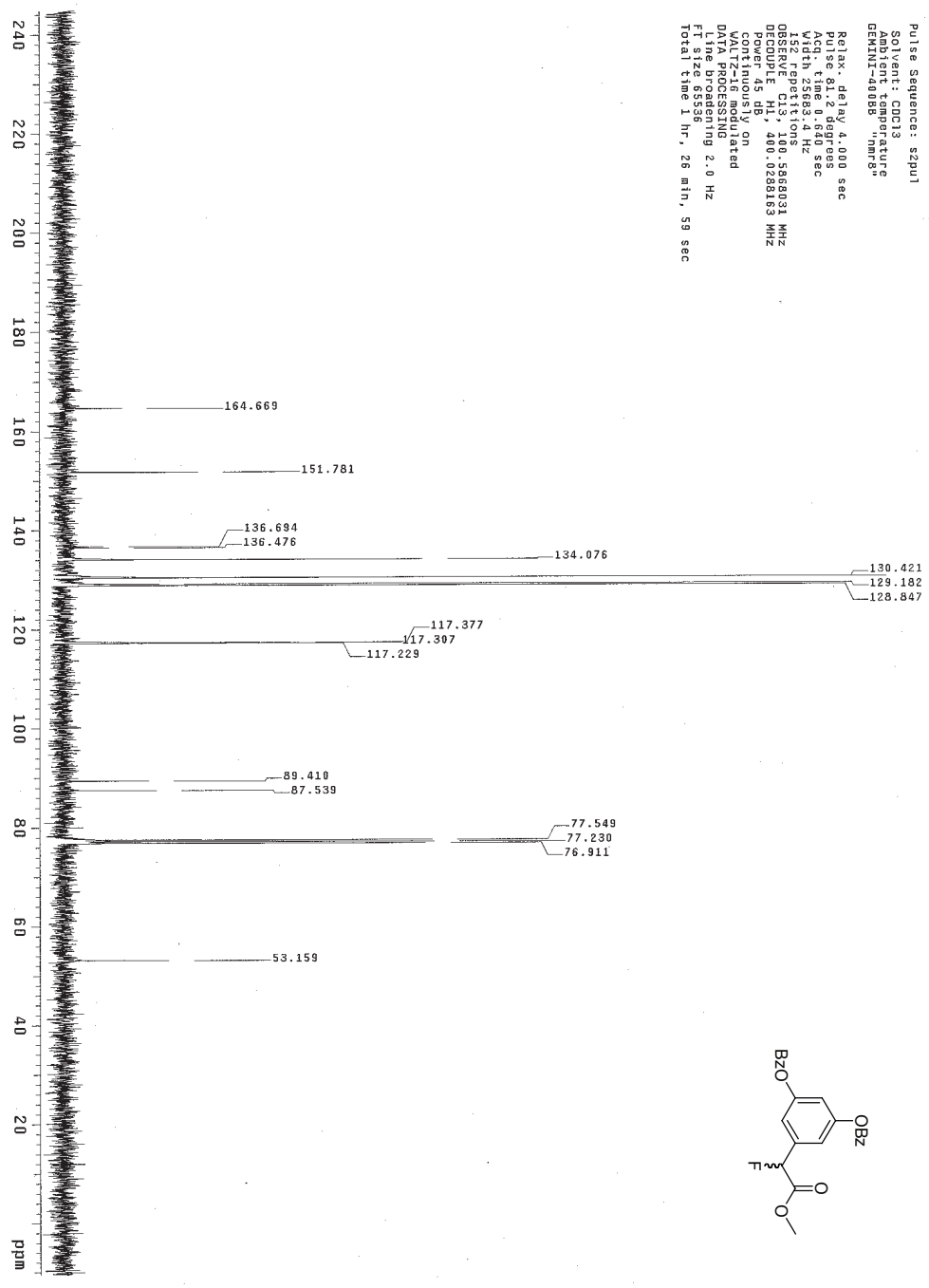

8.200
8.197
8.184
8.180
8.176
7.668
7.665
7.661
7.651
7.646
7.631
7.627
7.534
7.530
7.514
7.499
7.495
7.314
7.309
7.265
7.260
5.968
3.791
3.779
3.759
2.249
2.247
2.218
4.61
2.58
4.79
1.77
1.00
2.98
3.12
ppm
BzO
OBz
O
O
O
O

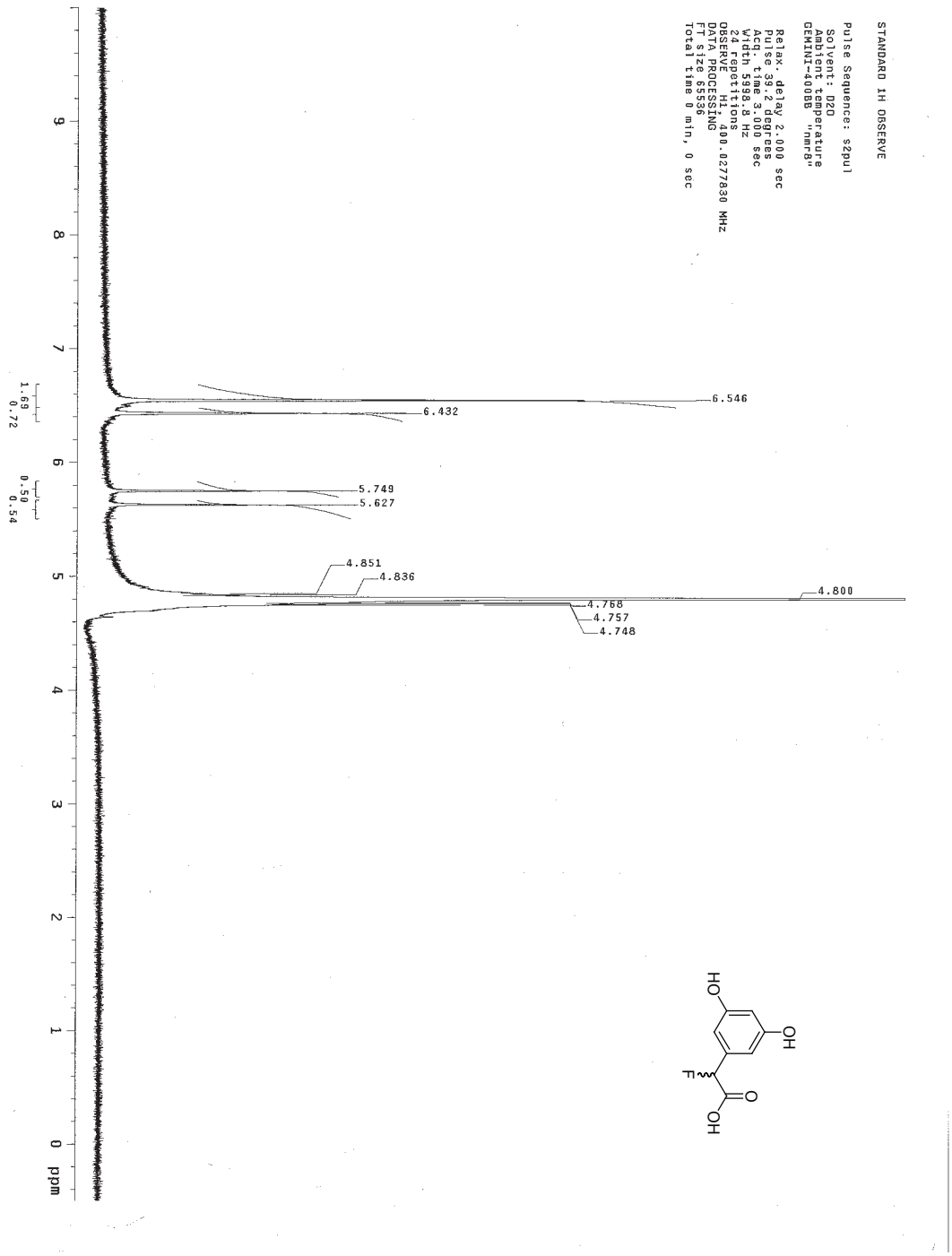

13C OBSERVE
Pulse Sequence: s2pul
Solvent: CDCl3
Ambient temperature
GEMINI-400BB "nmr8"
Relax. delay 4.000 sec
Pulse 81.2 degrees
Acq. time 0.640 sec
Width 25683.4 Hz
100 repetitions
OBSERVE C13, 100.5868077 MHz
DECOUPLE H1, 400.0288163 MHz
Power 45 dB
continuously on
WALTZ-16 modulated
DATA PROCESSING
Line broadening 2.0 Hz
FT size 65536
Total time 1 hr, 26 min, 59 sec
170.133
168.715
164.663
151.650
136.119
134.023
130.368
129.176
128.810
118.493
116.919
77.543
77.224
76.912
73.670
53.098
20.877
ppm
BzO
OBz

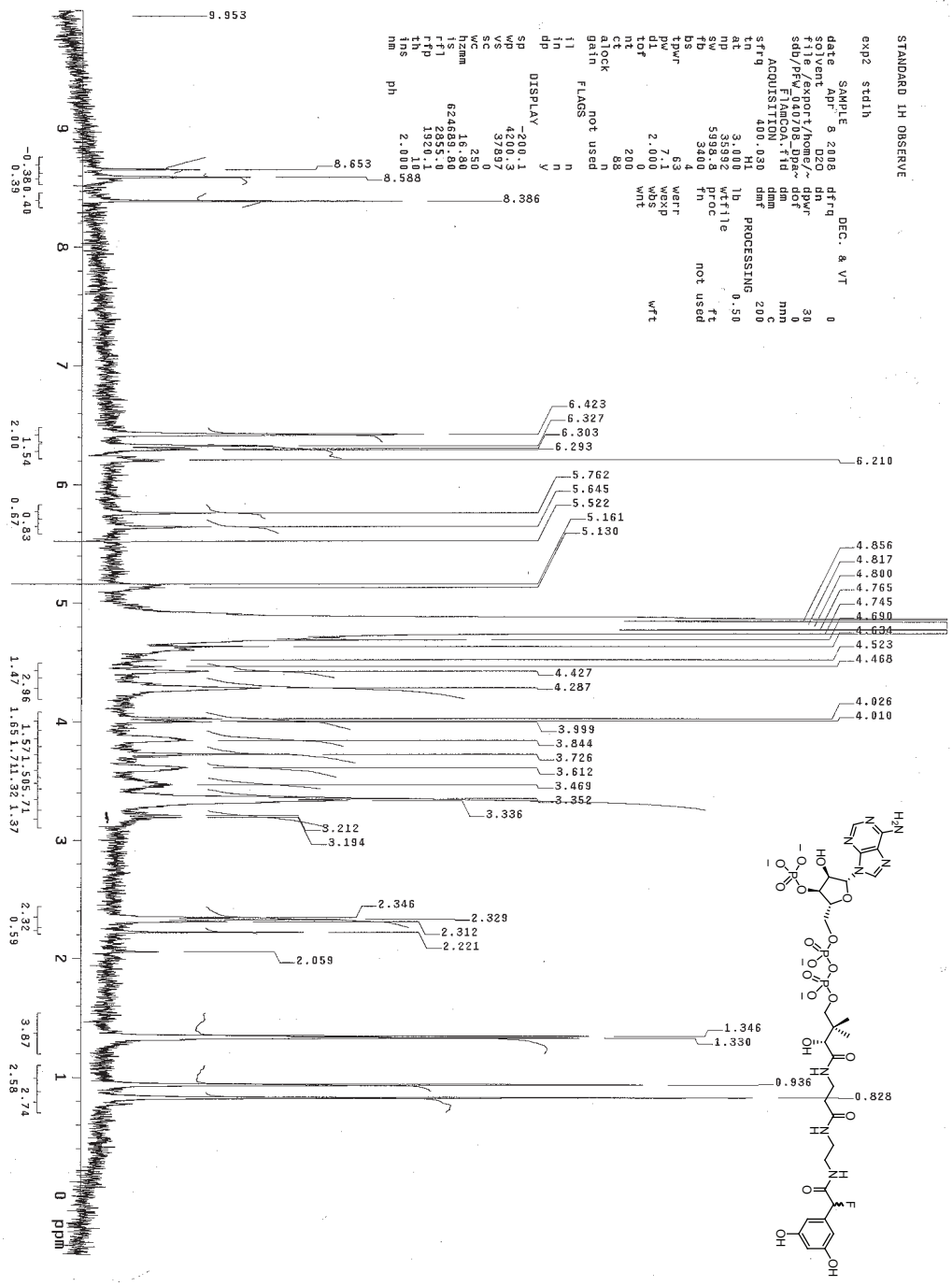

STANDARD 1H OBSERVE
Pulse Sequence: s2pul
Solvent: CDCl3
Ambient temperature
File: PFV_IV_061307H1
GEMINI-400BB "nmr8"
Relax. delay 2.000 sec
Pulse 39.2 degrees
Acq. time 3.000 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 400.0267918 MHz
DATA PROCESSING
FT size 65536
Total time 2 min, 59 sec
8.260
8.259
8.240
8.237
7.715
7.697
7.678
7.585
7.566
7.546
7.368
7.363
7.265
7.260
7.255
5.327
5.314
3.856
3.807
3.793
3.61
2.19
4.34
1.97
1.00
3.11
BzO
OBz
OH
O
O
ppm


13C OBSERVE
Pulse Sequence: s2pul
Solvent: CDCl3
Ambient temperature
GEMINI-400BB "nmrg"
Relax. delay 4.000 sec
Pulse 81.2 degrees
Acq. time 0.640 sec
Width 25683.4 Hz
100 repetitions
OBSERVE C13, 100.5868117 MHz
DECOUPLE H1, 400.0288163 MHz
Power 45 dB
continuously on
WALTZ-16 modulated
DATA PROCESSING
Line broadening 2.0 Hz
FT size 65536
Total time 1 hr, 26 min, 59 sec
173.280
164.732
151.516
140.855
133.912
130.312
129.260
128.746
117.478
115.927
77.549
77.230
76.911
72.196
53.479
ppm
BzO
OBz
OH
O
O


STANDARD 1H OBSERVE
Pulse Sequence: s2pul
Solvent: CDCl3
Ambient temperature
GEMINI-400BB "nmr8"
Relax. delay 2.000 sec
Pulse 39.2 degrees
Acq. time 3.000 sec
Width 5998.8 Hz
12 repetitions
OBSERVE H1, 400.0268153 MHz
DATA PROCESSING
FT size 65536
Total time 2 min, 59 sec
8.208
8.204
8.186
8.183
7.658
7.643
7.640
7.546
7.541
7.526
7.511
7.506
7.316
7.312
7.311
7.290
7.285
7.260
5.927
5.809
3.827
1.600
0.001
3.71
2.21
4.01
3.03
0.45
0.46
3.00
0.55
ppm
BzO
OBz
F
O
O


13C OBSERVE
Pulse Sequence: s2pul
Solvent: CDCl3
Ambient temperature
GEMINI-400BB "nmr8"
Relax. delay 4.000 sec
Pulse 81.2 degrees
Acq. time 0.640 sec
Width 25683.4 Hz
152 repetitions
OBSERVE C13, 100.5868031 MHz
DECOUPLE H1, 400.0288163 MHz
Power 45 dB
continuously on
WALTZ16 modulated
DATA PROCESSING
Line broadening 2.0 Hz
FT size 65536
Total time 1 hr, 26 min, 59 sec
164.669
151.781
136.694
136.476
134.076
130.421
129.182
128.847
117.377
117.307
117.229
89.410
87.539
77.549
77.230
76.911
53.159
BzO
OBz
Ph
O
O
240
220
200
180
160
140
120
100
80
60
40
20
ppm


STANDARD 1H OBSERVE
Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
GEMINI-400BB "nmr8"
Relax. delay 2.000 sec
Pulse 39.2 degrees
Acq. time 3.000 sec
Width 5998.8 Hz
24 repetitions
OBSERVE H1, 400.0277830 MHz
DATA PROCESSING
FT size 65536
Total time 0 min, 0 sec
6.546
6.432
5.749
5.627
4.851
4.836
4.800
4.768
4.757
4.748
1.69
0.72
0.50
0.54
ppm
HO
OH
F
O
OH


STANDARD 1H OBSERVE
exp2 std1h
SAMPLE
date Apr 8 2008
solvent D2O
file /export/home/~
sdb/pfw_1400/8_pp/~
ACQUISITION
sfrq 400.030
tn H1
at 3.000
np 35992
sw 5998.8
fb 3400
bs 4
tpwr 63
pw 7.1
pw1 2.000
d1 0
tof 200
nt 88
ct
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -200.1
wp 4200.3
vs 37897
sc 0
wc 250
hzmm 16.80
is 624689.80
rfl 2855.0
rfp 1920.1
th 2.000
ins
nm ph
dfrq
dn
dpwr
dof
dm
dmm
dmf
PROCESSING
lb 0.50
wtfile
proc ft
fn not used
wft
wexp
wbs
wnt
DEC. & VT
0
30
0
nnn
c
200
9.953
8.653
8.588
8.386
6.423
6.327
6.303
6.293
6.210
5.762
5.645
5.522
5.161
5.130
4.856
4.817
4.800
4.765
4.745
4.690
4.634
4.523
4.468
4.427
4.287
4.026
4.010
3.999
3.844
3.726
3.612
3.469
3.352
3.336
3.212
3.194
2.346
2.329
2.312
2.221
2.059
1.346
1.330
0.936
0.828
ppm

Chapter 5: Biochemical characterization and efforts toward structural characterization of a non-ribosomal peptide synthetase module from *Thermobifida fusca*.

Introduction

Nonribosomal peptide (NRP) natural products are structurally diverse small molecules and have a variety of biological functions. As secondary metabolites, nonribosomal peptides are used by bacteria and other organisms for many fundamental processes.[1] Vancomycin is an NRP with potent antibiotic function, evolved by bacteria to kill competing species in the same ecological niche and used by humans to treat bacterial infection.[2] Siderophore secondary metabolites are frequently NRP natural products, and used to scavenge ferric iron ($Fe(OH)_3$) through the formation of soluble chelation complexes.[3] NRP natural products can be adapted for a variety of biological activities; each of these activities can be achieved through different chemical functionalities.

Iron is fundamental to the most basic life processes.[4] In biology, reaction of molecular oxygen with iron is the foundation of metabolism. Although iron is abundant on the earth, a major drawback to its use in biological systems is its virtual insolubility in water (k_{sp} =10^{-18}M).[5] Siderophores are the bacterial adaptation to the selective pressure of acquiring an abundant yet inaccessible element.[6] Common iron

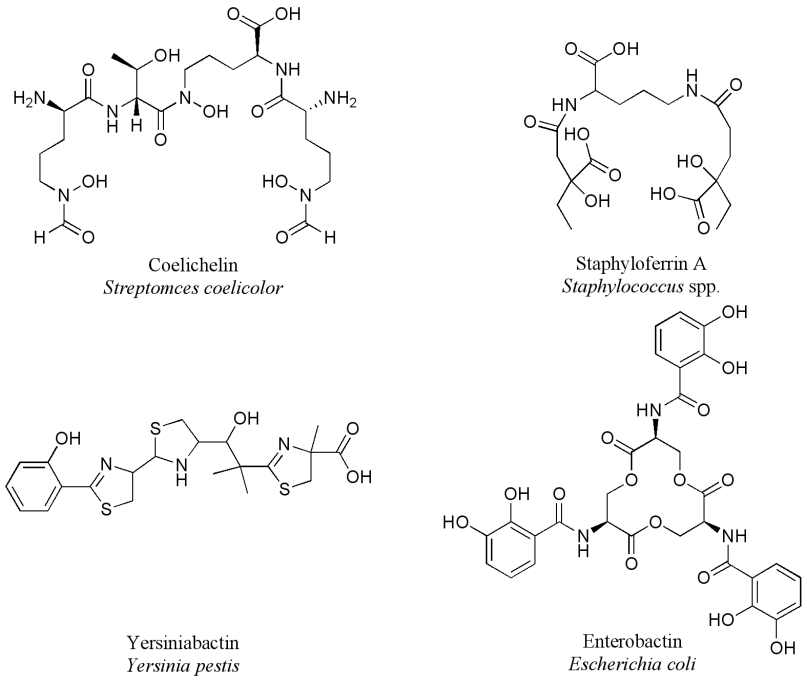
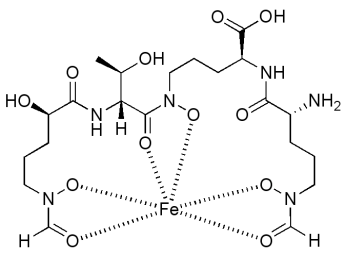


Figure 1: Siderophores containing a variety of iron binding motifs.

binding/chelating structural motifs in siderophores include hydroxamates, catechols, and α-hydroxyacids (Figure 1). Siderophores use these structural motifs to bind Fe^{III}, which forms a water-soluble complex between the siderophore and the otherwise insoluble Fe^{III} molecule (Figure 2). These complexes are then taken into the bacteria through transport proteins where the iron is released from the complex for use in the cell.

The identification of new siderophores, as well as many other natural products, often begins with the identification of orphan biosynthetic gene clusters.[7] Using homology searches with the primary amino acid sequence from an orphan gene cluster it is possible to identify the function of the gene products. Based on homology to known proteins, the function of an entire gene cluster can be hypothesized. When applied to a biosynthetic pathway, this approach can lead to accurate identification of a natural product's function based purely on the amino acid sequence of the proteins in the pathway.[8]



Coelichelin
Streptomces coelicolor

Figure 2: Coelichelin bound to iron

With the hypothesized function of a small molecule, assay-guided fractionation can be used to isolate that small molecule from fermentation of the producing organism. This process is called "genome mining" as it begins with analysis of a bacterial genome in order to find small molecule natural products.[9,10]

Genome mining is particularly applicable to NRP natural product discovery as assembly-line nonribosomal peptide synthetases (NRPS) are often co-linear to their products.[1] It is possible to identify the amino acid building blocks activated by NRPS systems based on the sequence of adenylation (A) domains in the assembly line.[11,12] Based on the sequence from the genome, it is possible to propose the structure of a

natural product without ever isolating it. However, there are several aspects of NRP biosynthesis that make the structure and even the amino acid sequence of these small molecules unpredictable. These aspects include uncommon amino acid incorporation and module skipping and repeating.[13] Therefore, in spite of the predictions made through bioinformatic techniques, it is necessary to experimentally characterize natural products and biochemically characterize the enzymes that assemble natural products. The structure of fuscachelin A demonstrates that the pathway to NRP natural products is difficult to predict based on bioinformatic analysis.

Siderophore gene cluster in *Thermobifida fusca*

An orphan gene cluster in *T. fusca* was found to contain genes corresponding to a NRPS biosynthetic pathway.[14] Three NRP synthetase genes designated *fscGHI* are continugous in the cluster and correspond to five peptide elongation modules. The first gene, *fscG*, codes for a 390 kDa protein which contains three extension modules (Figure 3). NRPS modules are made up of individually folded domains that are connected to each other via linker regions. This allows a single polypeptide chain to contain multiple adenylation (A), condensation (C) and peptidyl carrier protein (PCP) domains. FscG is a protein which contains three modules each made up of a C, A and

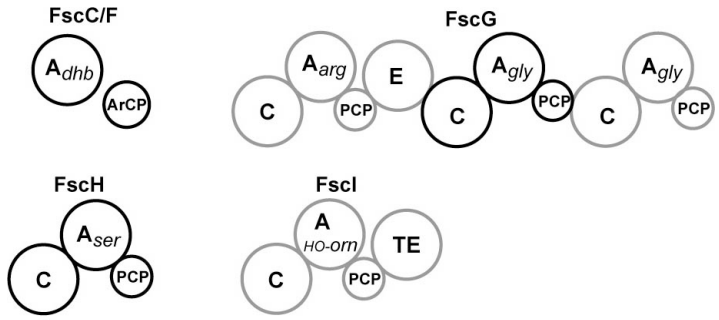


Figure 3: NRPS assembly line found in the *T. fusca* orphan gene cluster

PCP domains. This single protein therefore contains the function of six separate enzymes. FscH is a module containing a single condensation domain, adenylation domain and peptidyl carrier protein. FscI is a termination module with a C, A and PCP domain, and also contains a thioesterase (TE) domain. TE domains are responsible for cleaving the thioester linkage between the peptide natural product and the NRPS assembly line, and can also have cyclization functionality. The amino acids activated by these modules can be predicted based on their A domain sequence (Figure 4).[11,12] However the indication that this biosynthetic cluster encodes a siderophore came from three genes upstream from the *fscGHI* NRPS genes.

One of the common structural motifs found in siderophores are 2,3-dihydroxybenzoate (2,3-Dhb) groups. The proteins FscA, FscB and FscD are homologous to the well characterized catecholate biosynthetic enzymes: isochorismate synthase, isochorismatase and 2,3-dihydro-Dhb dehydrogenase.[15,16,17] FscC, an A domain

FscC	**P M P A Q G V V**	
	P L P A Q G V V	(2,3-Dhb, bacillibactin)
$FscG_1$	**D A D D S G C V**	
	D V W N F G F V	(L-Arg, nodularin)
	D A E D L G F V	(L-Arg, vanchrobactin)
	D V A D V G A I	(L-Arg, syringomycin)
$FscG_2$	**D I L Q F G V I**	
	D I L Q L G L I	(gly, nostopeptolide A)
$FscG_3$	**D I L Q V G V I**	
	D I L G L G L I	(gly, nostopeptolide A)
FscH	**D V W H I S L V**	
	D V W H I S L V	(L-Ser, nostopeptolide A)
FscI	**D M E N L G L I**	
	D M E N L G L I	(L-HO-Orn, coelichelin)

Figure 4: Specificity of each A domain in the Fsc gene cluster. Specificity of FscI could not be determined before the structure of fuscachelin A was solved.

with predicted specificity for Dhb, and a dedicated aryl-carrier protein, FscF, are present as stand-alone domains for incorporation of Dhb as the initial building block.

Based on the presence of the five NRPS extension modules and FscC/F, along with the specificity-conferring code of the adenylation domains, a structure for the product of the *T. fusca* orphan biosynthetic pathway can be proposed: a pentapeptide, N-capped with Dhb (Figure 5). This prediction represents a structure unlike any other characterized siderophore. However, the actual structure of the molecule as determined by NMR and mass spectrometry techniques is quite different.

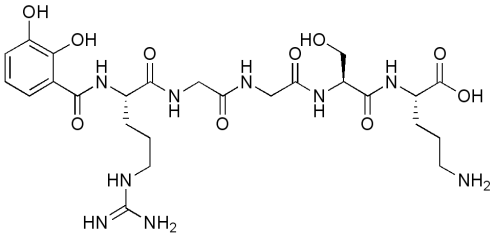


Figure 5: Predicted structure of fuscachelin based on the gene cluster sequence

Structure elucidation of the *T. fusca* siderophores, the Fuscachelins

Eric Dimise performed structural characterization on the fuscachelin siderophores.[18] The small molecules were isolated from *T. fusca* preps grown in iron-depleted Hagerdal media.[19] Siderophore activity was monitored by the chrome

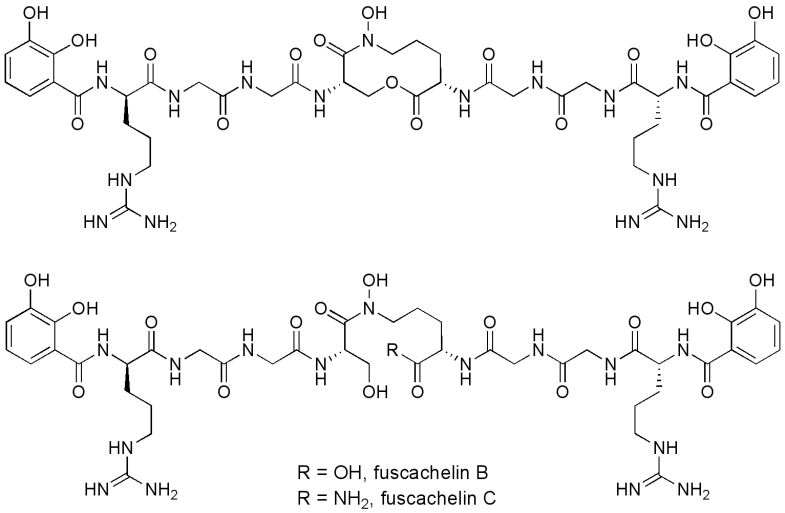


Figure 6: Structures of fuscachelin A/B/C as determined by NMR and mass spectrometry.

azurol S (CAS) assay in order to determine the presence of an iron-binding molecule.[20] The CAS assay allowed for identification of siderophore containing fractions during purification. Once pure, the fuscachelin molecules were structurally characterized through a series of NMR and mass spectrometry techniques (Figure 7).[18] Fuscachelin A is a heterodimer with a macrolactone at the center of the natural product. Fuscachelin B and C were also isolated from *T. fusca*, which also had with siderophore activity, and are believed to be degradation products of fuscachelin A. The structural characterization of fuscachelin A allowed for the biochemical characterization of the terminal NRPS module FscI.

Biochemical characterization of the NRPS module FscI

In order to corroborate that the orphan biosynthetic gene cluster identified in *T. fusca* actually produces fuscahcelin A, FcsI was biochemically characterized. The 4.0 kb gene for FscI was cloned from *T. fusca* genomic DNA into the pET30a expression vector. The four-domain protein was then overexpressed in *E. coli* to high levels and was purified to homogeneity as judged by SDS/PAGE analysis. A pyrophosphate exchange

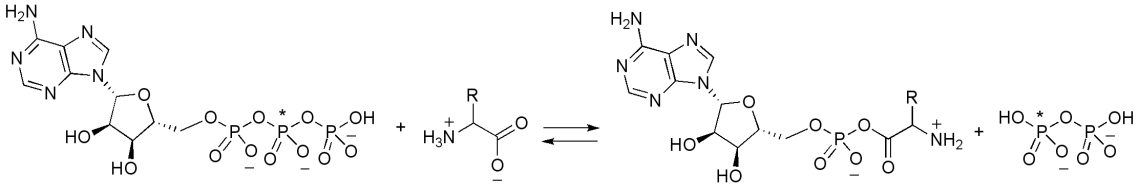


Figure 7: Pyrophosphate exchange assay: the reaction catalyzed by adenylation domains in NRPS modules. (*) indicates radioactive ^{32}P

assay of the FscI adenylation domain was used to determine the module's amino acid specificity.[21] The pyrophosphate A domain assay screened the twenty canonical amino acids as well as ornithine and L-*N*-ε–hydroxyorninithine (L-*HO*Orn) (Figure 8). The

results of the A domain assay show FscI has a significant preference for L-*HO*Orn (Figure 9). This implies the flavin monooxygenase FscE hydroxylates the free amino acid L-ornithine, not the mature peptide. Prior to the structure solution of the fuscachelin molecules, the specificity of FscI could not be determined, as L-*HO*Orn was not considered to be a candidate for activation by FscI.

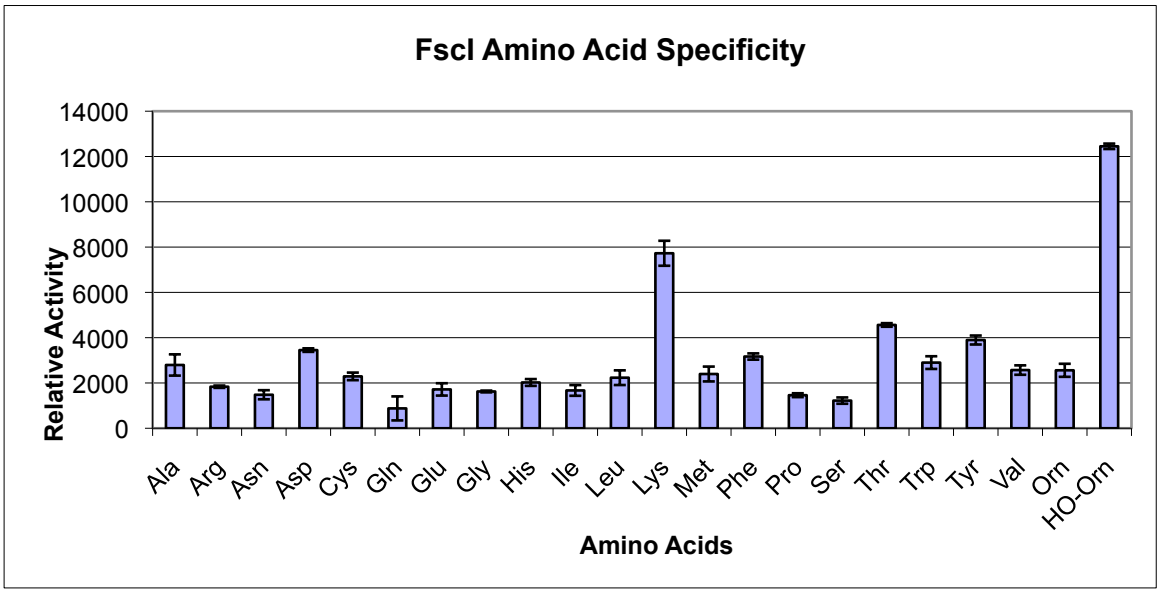


Figure 8: Relative activity of FscI with various amino acids as judged by the pyrophosphate exchange assay.

Proposed biosynthetic pathway of fuscachelin A

With the structure of the natural product and biochemical evidence linking its production to the identified gene cluster, it is possible to propose a biosynthetic pathway for the molecule (Figure 10). Fuscachelin A is most likely the product of the gene cluster as degradation via hydrolysis under mildly basic conditions leads to formation of fuscachelin B. Fuscachelin A is a heterodimer peptide composed of a tetrapeptide and

pentapeptide, each capped on the N-terminus with a Dhb group. The unusual chemistry of the pathway occurs at the condensation domain of FscI. Once FscH has coupled serine to the peptide chain, FscI couples the tetrapeptide to the ε-nitrogen of L-*HO*Orn. At this

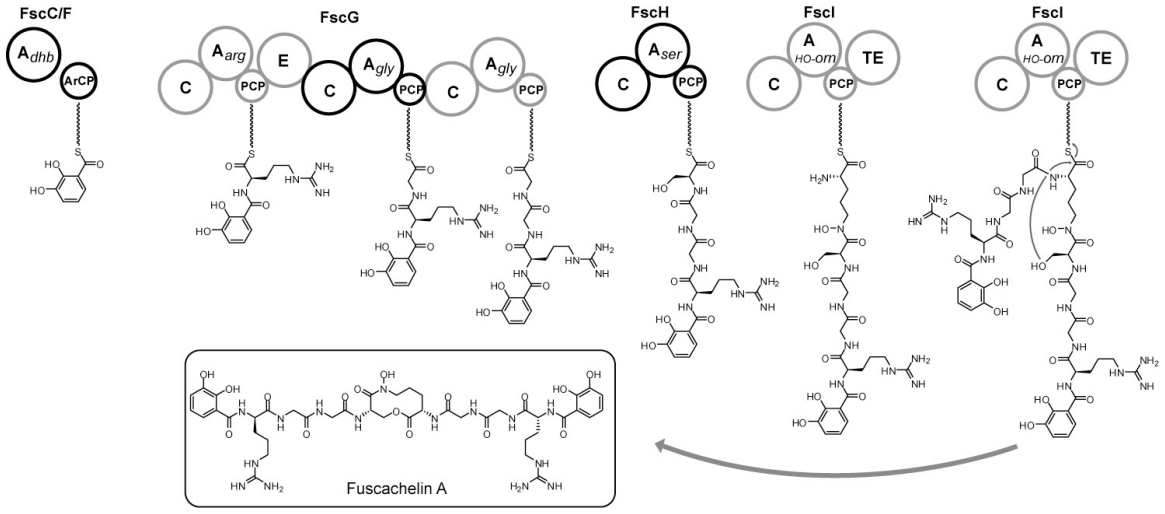


Figure 9: Proposed biosynthetic pathway to fuscachelin A

point we suggest a second Dhb-capped tripeptide from FscG is coupled to L-*HO*Orn, this time at the α-nitrogen. In each coupling, the stand-alone modules of FscH or FscI must "dock" with FscG. A recognition sequence is required to facilitate the docking and orientation between the NRPS modules. Therefore, FscG must contain a region that can recognize FscH and FscI and allow for coupling between thioester-tethered peptides from both stand-alone modules. After the tandem coupling to the α- and ε-nitrogens of L-*HO*Orn, the thioesterase domain of FscI performs a macrocyclization to form the 10-membered depsipeptide ring of fuscachelin A. The FscI thioesterase domain shows homology to the TE domain of DhbF from bacillibactin biosynthesis, which catalyzes a similar reaction.[22] With the variety of natural products synthesized by NRPS modules, structural characterization of a multi-domain module would provide great insight into the chemistry of these systems.

Efforts towards structural characterization of FscI

The NRPS enzyme FscI is an ideal choice for protein crystallography as it originates from an organism that is a moderate thermophile. Thermophilic bacteria provide proteins that are often more amenable to protein crystallography. Possible explanations for enhanced thermophilic protein crystallography are increased stability and reduced mobility at room temperature. This makes all NRPS modules from the fuscachelin pathway good candidates for crystallography.

Several efforts towards solving the macromolecular crystal structure of FscI were attempted. Deborah Mitchell was successful in her attempts to crystallize FscI, however she could not reproduce the crystals. We were able to collect a full data set on an FscI crystal that diffracted to 3.0 Å. Phase information for FscI was sought after using molecular replacement techniques. Crystal structures of condensation, adenylation, peptidyl carrier protein and thioesterase domains have been previously solved.[23,24,25,26] Therefore homology models of the four domains that compose FscI are available for molecular replacement. In the case of a multi-domain enzyme, solving a structure with molecule replacement can be difficult. Initial attempts towards a solution with molecular replacement were unsuccessful.

The next strategy used to solve the structure of FscI was crystallization of a seleno-methionine (SeMet) derivative. SeMet FscI would not crystallize under the original native conditions, or when screened against the Hampton Research High-Throughput Screen, this lead to another effort of molecular replacement.

The program PHASER provided some hope that the structure of FscI could be solved.[27] Using the original data set from Deborah Mitchell's crystal, multiple domains

were used as search models in a single search. This is made possible by multiple "ensembles" used as search models in PHASER. An ensemble is simply a protein structure, or multiple homologous protein structures overlaid on each other. Initial results from the model searches with PHASER gave promising results. Statistics were high enough to indicate a solution had been achieved. However, no solutions from PHASER provided phase information for the FscI data set.

The structure of a terminal NRPS domain was recently published, which contained four domains homologous to FscI.[28] Current and future efforts will use this structure as a search model to find a phase solution for FscI.

Conclusion

Fuscachelin A was discovered through a genome mining approach. The proposed biosynthetic pathway contains unusual aspects that demonstrate the flexibility of NRP assembly line chemistry. The biochemical characterization of FscI shows the importance of experimental evidence for determining enzyme substrate specificity. The power of bioinformatic predictions will increase with experimental characterization of biological systems.

Materials and Methods

Cloning, Expression, and Purification of FscI.

The gene for *fscI* was amplified by using PCR from *T. fusca* genomic DNA with the following primers: 5′-GCG GAA TTC ACC ACC GCA GCC GCG GGT (EcoRI), 5′-GCG AAG CTT CTA GCT GTG TCC GGA TCG (HindIII). The PCR products were purified through agarose gel electrophoresis and gel extraction (Qiagen) and cleaved with the EcoRI and HindIII restriction endonucleases. The *fscI* gene was then ligated into the plasmid pET30a. The plasmid was transformed into *E. coli* BL21(DE3) cells for gene expression. Cultures (1 L) were grown to A600 = 0.5–0.7 at 37 °C, at which point the shaker was cooled to 18 °C, and overexpression was initiated by the addition of 50 μM IPTG. Cultures were continued for 18 h and were harvested by centrifugation, followed by resuspension in 500 mM NaCl, 20 mM Tris·HCl (pH 7.5) and lysed by passage through a French pressure cell at 1,000 psi. Lysate was centrifuged at 10,000 rpm for 20 min in a Beckman Coulter J2-HS centrifuge. The supernatant was incubated for 1 h with 1 ml of metal-affinity resin (Talon resin; Clontech). Resin was washed with 4 × 10 ml of 500 mM NaCl, 20 mM Tris·HCl (pH 7.5), and protein was eluted with 2 × 10 ml of 500 mM NaCl, 20 mM Tris·HCl (pH 7.5), 250 mM imidazole. Protein was dialyzed against 100 mM NaCl, 20 mM Tris·HCl (pH 7.5), 1 mM β-mercaptoethanol, 10% (vol/vol) glycerol, concentrated to 17 μM, and flash frozen.

Pyrophosphate Exchange Assay.

Amino acid-dependent ATP-sodium pyrophosphate assays were performed as follows. A 100 µl reaction contained 75 mM Tris·HCl (pH 8.0), 10 mM MgCl2, 5 mM DTT, 5 mM ATP, 1 mM Na432P2O7, 100 µg/mL BSA, 1 mM amino acid, 2 µM FscI. Reactions were initiated by addition of enzyme and incubated at 30 °C for 0.5 h. The reaction was quenched by the addition of 500 µl of 3.5% charcoal, 1.6% perchloric acid, 200 mM $Na_4P_2O_7$. The charcoal was centrifuged and resuspended twice with 500 µl of 1.6% perchloric acid, 200 mM Na4P2O7. After washing, the charcoal was mixed with 3 ml of scintillation fluid and read by a Beckman–Coulter LS 6500 scintillation counter. All reactions were performed in triplicate.

References:

1 Fischbach, M. A. and Walsh, C. T., Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106** (8), 3468 (2006).

2 Hubbard, B. K. and Walsh, C. T., Vancomycin assembly: nature's way. *Angew Chem Int Ed Engl* **42** (7), 730 (2003).

3 Fischbach, M. A., Lin, H., Liu, D. R., and Walsh, C. T., How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nat Chem Biol* **2** (3), 132 (2006).

4 A. Sigel, H.E. Sigel, *Metal Ions in Biological Systems: Iron Transport and Sotrage in Microorganisms, Plants, and Animals*. (Dekker, New York, 1998).

5 Sigel, A. and Sigel, H. (Eds.), *Metal Ions in Biological Systems: Iron Transport and Storage in Microorganisms, Plants, and Animals*. (Marcell Dekker Inc., New York, 1998).

6 Schaible, U. E. and Kaufmann, S. H., Iron and microbial infection. *Nat Rev Microbiol* **2** (12), 946 (2004).

7 Challis, G. L., Genome mining for novel natural product discovery. *J Med Chem* **51** (9), 2618 (2008).

8 Gross, H., Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl Microbiol Biotechnol* **75** (2), 267 (2007).

9 Challis, G. L., Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* **154** (Pt 6), 1555 (2008).

10 Wilkinson, B. and Micklefield, J., Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol* **3** (7), 379 (2007).

11 Stachelhaus, T., Mootz, H. D., and Marahiel, M. A., The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6** (8), 493 (1999).

12 Challis, G. L., Ravel, J., and Townsend, C. A., Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7** (3), 211 (2000).

13 Lautru, S., Deeth, R. J., Bailey, L. M., and Challis, G. L., Discovery of a new peptide natural product by Streptomyces coelicolor genome mining. *Nat Chem Biol* **1** (5), 265 (2005).

14 Lykidis, A. et al., Genome sequence and analysis of the soil cellulolytic actinomycete Thermobifida fusca YX. *J Bacteriol* **189** (6), 2477 (2007).

15 Liu, J., Duncan, K., and Walsh, C. T., Nucleotide sequence of a cluster of Escherichia coli enterobactin biosynthesis genes: identification of entA and purification of its product 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase. *J Bacteriol* **171** (2), 791 (1989).

16 Liu, J., Quinn, N., Berchtold, G. A., and Walsh, C. T., Overexpression, purification, and characterization of isochorismate synthase (EntC), the first enzyme involved in the biosynthesis of enterobactin from chorismate. *Biochemistry* **29** (6), 1417 (1990).

17 Rusnak, F. et al., Subcloning of the enterobactin biosynthetic gene entB: expression, purification, characterization, and substrate specificity of isochorismatase. *Biochemistry* **29** (6), 1425 (1990).

18 Dimise, E. J., Widboom, P. F., and Bruner, S. D., Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile Thermobifida fusca. *Proc Natl Acad Sci U S A* **105** (40), 15311 (2008).

19 Hagerdal, B. G., Ferchak, J. D., and Pye, E. K., Cellulolytic Enzyme System of Thermoactinomyces sp. Grown on Microcrystalline Cellulose. *Appl Environ Microbiol* **36** (4), 606 (1978).

20 Schwyn, B. and Neilands, J. B., Universal chemical assay for the detection and determination of siderophores. *Anal Biochem* **160** (1), 47 (1987).

21 Tang, G. L., Cheng, Y. Q., and Shen, B., Chain initiation in the leinamycin-producing hybrid non-ribosomal peptide/polyketide synthetase from Streptomyces atroolivaceus S-140: discrete, monofunctional adenylation enzyme and peptidyl carrier protein that directly load (D)-alanine. *J Biol Chem* (2007).

22 May, J. J., Wendrich, T. M., and Marahiel, M. A., The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J Biol Chem* **276** (10), 7209 (2001).

23 Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P., Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo J* **16** (14), 4174 (1997).

24 Couch, R. et al., Characterization of CmaA, an adenylation-thiolation didomain enzyme involved in the biosynthesis of coronatine. *J Bacteriol* **186** (1), 35 (2004).

25 May, J. J., Kessler, N., Marahiel, M. A., and Stubbs, M. T., Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A* **99** (19), 12120 (2002).

26 Weber, T. et al., Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure Fold Des* **8** (4), 407 (2000).

27 COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4., The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D* **50**, 760 (1994).

28 Tanovic, A., Samel, S. A., Essen, L. O., and Marahiel, M. A., Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321** (5889), 659 (2008).

Appendix: Complex oxidation chemistry in the biosynthetic pathways to vancomycin antibiotics.

Adapted from an invited review by P.F. Widboom and S.D. Bruner to be published in Chembiochem

Introduction

Studying natural product biosynthesis often leads to the discovery of unique and powerful enzymes behind the biological assembly of small molecules.[1] Several interesting oxidation enzymes are found in the vancomycin biosynthetic pathway. The study of these enzymes through genetic, biochemical and X-ray crystallographic analysis has illuminated their role in the biosynthesis of vancomycin, and the mechanistic details of powerful oxidation chemistry. In this appendix, three diverse examples of oxidation chemistry involved in different aspects of vancomycin biosynthesis are discussed.

A growing understanding of all aspects of vancomycin biosynthesis expands our knowledge of powerful chemistry found in nature. It also leads to the manipulation these systems towards the goal of combinatorial biosynthesis, which will allow for the facile creation of potential analogs to overcome resistance.[2]

The non-heme iron dioxygenase, HmaS.

Natural products of the vancomycin family contain the amino acid L-4-hydroxyphenylglycine (Hpg), which is always a partner in aryl/aryl and aryl ether crosslinks found in the mature natural products.[1] Isotope labelling experiments provided the first Insights into the biosynthetic pathway to Hpg by establishing that Hpg is derived from L-tyrosine as shown in Figure 2A.[3] The first enzyme in the pathway to 4-Hpg is hydroxymandelic acid synthase (HmaS),

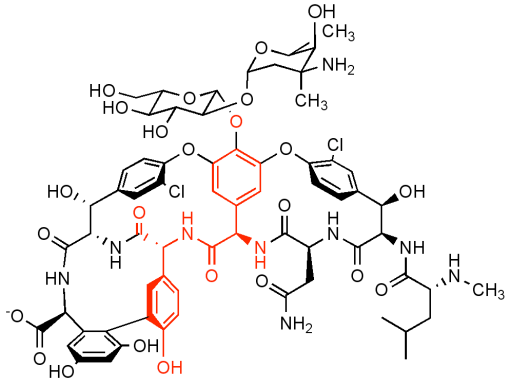


Figure 1: Vancomycin with 4-Hpg highlighted

which performs an oxidative decarboxylation/hydroxylation of the substrate 4-hydroxyphenylpyruvate, an intermediate in the shikimic acid primary metabolic pathway. Several gene clusters in the vancomycin family contain a gene for prephenate dehydrogenase responsible for the conversion of prepenate to 4-hydroxyphenylpyruvate initiating the cyclic pathway to Hpg.[1] The product of HmaS, the α-hydroxyacid **2**, is elaborated to the amino acid through common enzyme chemistry: oxidization to an α-ketoacid **3** by a flavin-dependent enzyme then installation of the α-amine by the PLP-

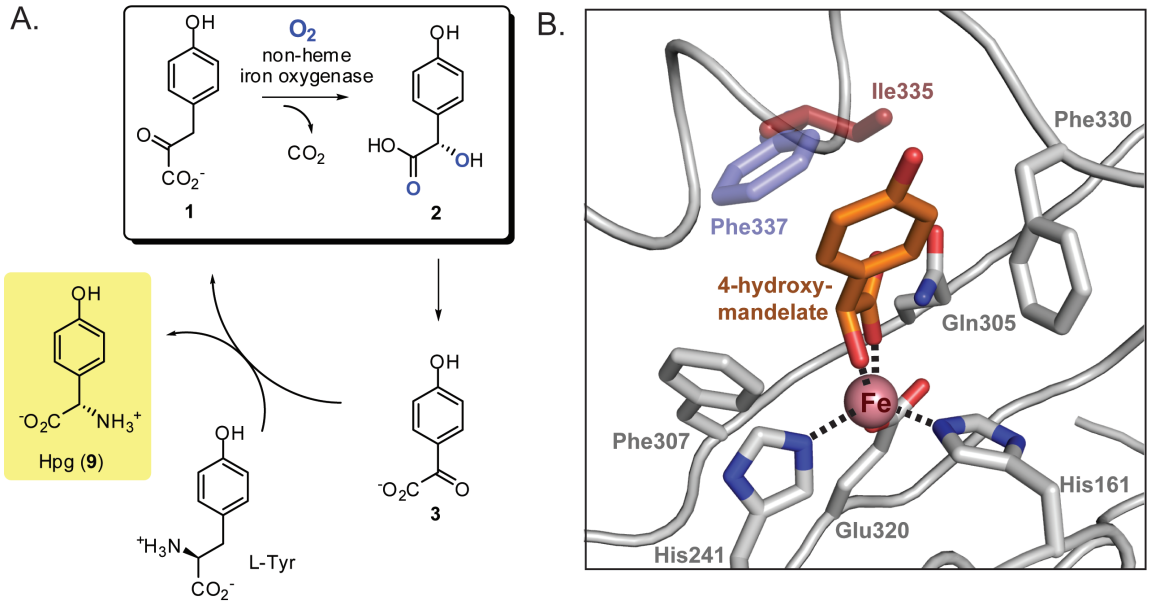


Figure 2: A. Catalytic cycle of 4-Hpg biosynthesis. B. Active site of HmaS bound to 4-hydroxyphenylpyruvate

dependant aminotransferase, HpgT. Interestingly, HpgT processes **3** to the phenylglycine 4-Hpg using L-Tyr as the ammonia donor and in the process regenerates the starting α-keto acid **1**. HpgT is same enzyme the converts DpgX to Dpg in the 3,5-dihydroxyphenylglycine biosynthetic pathway.

The key step in the pathway is catalyzed by the mononuclear non-heme iron dioxygenase, HmaS. Most characterized non-heme iron oxygenases utilize α-

ketoglutarate as a cofactor to supply reducing equivalents leading to a high-energy iron (oxo) species. For HmaS, this role is served by the substrate, leading to loss of carbon dioxide and incorporation of two oxygen atoms into the product.[4] In addition, the hydroxylation step is unique to HmaS as the hydroxylation is selectively at the benzyl position as opposed to other known 4-hydroxyphenylpyruvate dioxygenases, which hydroxylates the phenyl ring.[5] Two groups described the selectivity and reactivity of HmaS concurrently.[6,7] Prior to this work, none of the intermediates along the pathway to Hpg had been identified. An examination of the biosynthetic gene cluster for the vancomycin homolog chloroeremomycin revealed one gene product with high sequence homology to p-hydroxyphenylpyruvate dioxygenase (HPPD). HPPD is an α-keto acid-dependent non-heme iron-dependent dioxygenase that catalyzes the decarboxylation/ hydroxylation at the aromatic ring of the phenylpyruvate substrate **1**. Biochemical

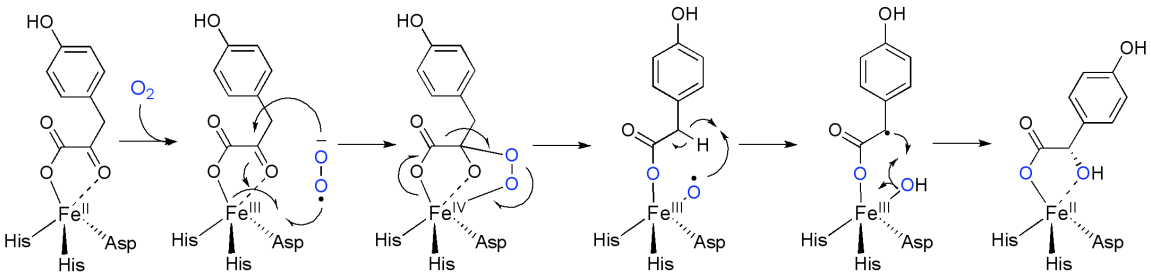


Figure 3: Proposed mechanism of HmaS

experiments confirmed that the chemistry of HmaS proceeds with novel regioselectivity for a p-phenylpyruvate dioxygenase forming 4-hydroxymandelate (**2**). The chemistry of HmaS is predicted to proceed analogously to other non-heme iron enzymes and is summarized in Figure 3. The substrate binds as a bidentate complex with the enzyme chelated iron. O_2 then coordinates to the iron and adds into the α-keto group of the substrate. A rearrangement results in loss of carbon dioxide and generation of a reactive

iron-oxo species. This high-energy intermediate is capable of abstracting a hydrogen atom from the benzyllic position of the substrate.

The Solomon group explored this difference in reactivity through spectroscopic and computational studies.[8] This study showed that HPPD and HmaS have similar substrate bound complexes. Spectroscopic data show comparable electronic structures for the active site and the iron-oxo intermediates. This leaves the difference in hydroxylation selectivity to the orientation of the substrate's aromatic ring in the active site. According to structural data, two phenylalanine residues, Phe337 and Phe341, in HPPD play a key role in π-stacking with the substrate.[9,10] Mutation Phe337 to Ile in HPPD confers some HmaS function, confirming the importance of substrate phenyl ring orientation in hydroxylation specificity.[11]

Brownlee *et al*. recently solved the crystal structure of HmaS bound to its product p-hydroxymandelate.[12] The crystal structure of HmaS shows the product chelated to the iron center through the benzylic alcohol and the carboxyl group. The shape of the hydrophobic pocket found in HmaS and HPPD dictates reaction specificity. HPPD has a larger hydrophobic pocket allowing differently shaped structures to evolve along the reaction pathway. The tighter contour of the HmaS structure does not allow this movement and restricts hydroxylation to the benzyl position, creating a product similar in shape to the substrate.

The biosynthesis of β-hydroxytyrosine

Tyrosine residues hydroxylated at the β-position are key building blocks of the vancomycin family members. These residues are partners in aryl-ether and aryl-aryl crosslinks with Hpg. In a subset of the vancomycin family the hydroxyl acts as scaffold for the attachment of sugar moieties. The chemistry of the hydroxylation has not been reconstituted biochemically but can be inferred from well-characterized systems.[13] The pathway involves three enzymes with one enzyme containing a distinct carrier domain.[14] This path represents a considerable amount of protein-based machinery to accomplish an overall insertion of one oxygen atom into L-Tyr. The first step involves activation and loading of L-Tyr by CepK (gene nomenclature is from the chloroeremomycin pathway), a didomain protein containing an adenylation domain that uses ATP to activate L-Tyr and load it as a phosphopantethienyl thioester on to the PCP carrier domain (Figure 4). The hydroxylation then occurs on the CepK-bound L-Tyr by an iron-heme monooxygenase of the P450 family. The chemical mechanism of the CH activation and hydroxylation for the P450 family of enzyme is well established and involves an

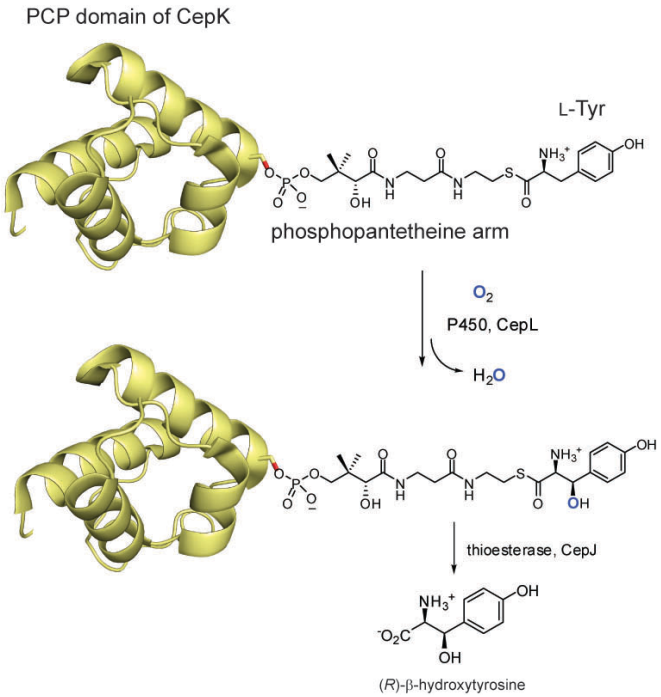


Figure 4: Chemistry catalyzed by CepL, the substrate is tethered to a CepK, an adenylation/PCP didomain via a phosphopantetheine linker

intermediate iron (IV) oxo species as detailed with camphor oxidase. This results in an enzyme-bound (*R*)-β-hydroxy-(*S*)-tyrosine linked as a thioester. To generate the product amino acid, a dedicated thioesterase (CepJ) hydrolytically cleaves the intermediate.

Aryl-aryl and aryl ether crosslinking enzymes

The oxidative crosslinks between aromatic residues on vancomycin create the rigid cup-like shape that gives the antibiotic its potency. Understanding the order of cyclization steps and substrate specificity of the cross-linking enzymes is key toward understanding the development of vancomycin analogues.

The order of aryl cyclization was first elucidated through analysis of *A. Mediterranei* knockout strains.[15] Upon fermentation of an *oxyA* knockout construct, a monocyclic heptapeptide vancomycin precursor was isolated. This showed the ability of either oxyB or oxyC to cyclize the C-O-D ring system. Mutant replacement strains of *A. Mediterranei* knocking out *oxyB* and *oxyC* were also created and their culture filtrates analyzed to determine the order of cyclization steps.[16] The *oxyB* knockout strain showed no cyclization product, implying OxyB as the initial cyclization enzyme (Figure 5). As discussed earlier the *oxyA* knockout gave a monocyclized product, implying it to be the second cyclizing enzyme.

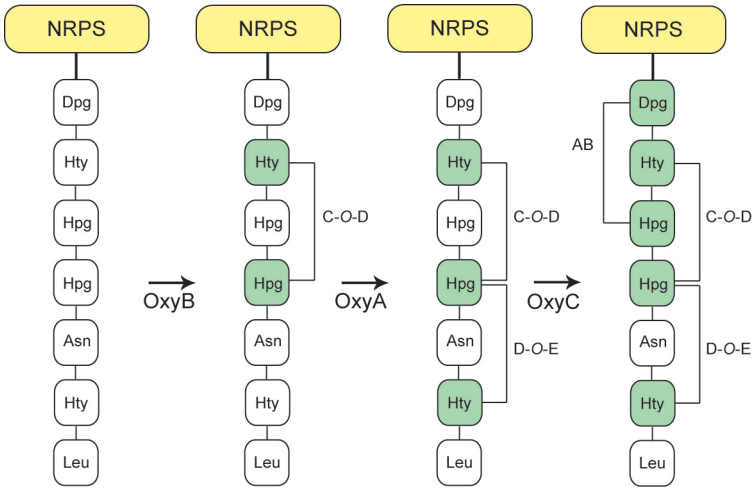


Figure 5: Schematic of the order of cyclization steps by OxyA/B/C

The *oxyC* knockout strain gave a bicyclized product. This work elucidated the order of oxidative crosslinking steps on the linear peptide vancomycin precursor, also showing the specificity of each cross-linking enzyme for a substrate with the proper degree of cyclization.

X-ray crystal structures of OxyB and OxyC were solved in the hope of elucidating the mechanism of this important step in vancomycin biosynthesis.[17,18] Both structures revealed a typical cytochrome P450 fold. However a striking difference between OxyB/C and other P450 enzymes is an exposed active site. This open conformation might allow for a hexa- or heptapeptide substrate tethered to a peptidyl carrier protein to have access to the heme center (Figure 6). When assays were attempted with purified oxyB/C in vitro with the linear heptapeptide as a substrate no activity was observed. Incubation of the enzymes with substrate under crystallization conditions yielded diffraction quality crystals however, electron density for the substrate was never observed. This implied the necessity of the peptide substrate to be tethered to a PCP domain on the NRPS assembly line in order for the formation of aryl cross links to occur.

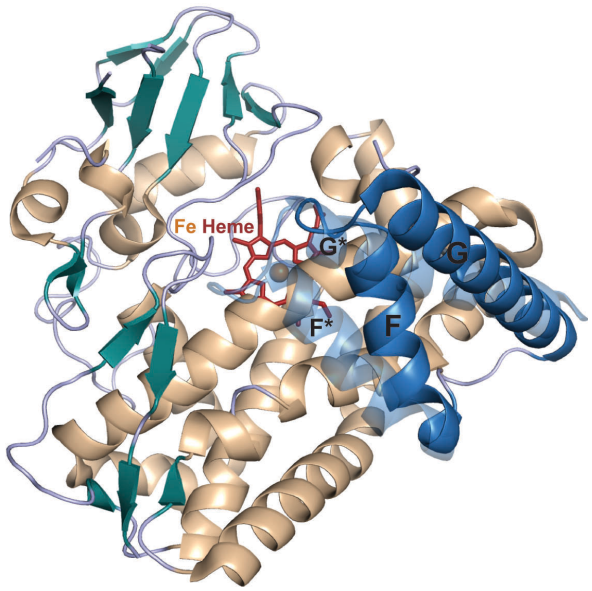


Figure 6: Crystal Structure of OxyB with P450nor. Shown in blue are helices F and G from Oxy B and in semi-transparent blue F and G from P450nor. Helices F and G of OxyB are shifted relative to a common P450 fold to allow a large substrate access to the active site. the cross-coupling enzyme.

Loading both the hepta and hexa peptide linear precursors onto an eighty-residue PCP domain facilitated in vitro cross-linking by OxyB.[19,20] This result indicated that *in vivo* coupling chemistry of OxyA/B/C occurs when the linear peptide precursor is still tethered to the NRPS assembly line through a thioester linkage (Figure 8).

To corroborate the conclusion that cross-linking occurs on the NRPS assembly line, deletion mutants of *A. balhimycina* that stall peptide formation at two different locations were studied.[21] The *dpgA* knockout mutant lacked the ability to synthesize the final amino acid building block, Dpg. The *bpsC* knockout mutant did not have a condensation domain for Dpg. In spite of an incomplete peptide being produced in both mutant strains, cross-linking activity was still observed. Most notable is the observation of a complete heptapeptide from the *dpgA* mutant that had undergone all three oxidative cyclization reactions. As Dpg was not available in the *dpgA* mutant, the final position in the heptapeptide had been substituted with Hpg. This shows a remarkable flexibility in the NRPS system, and in the oxidative cross-

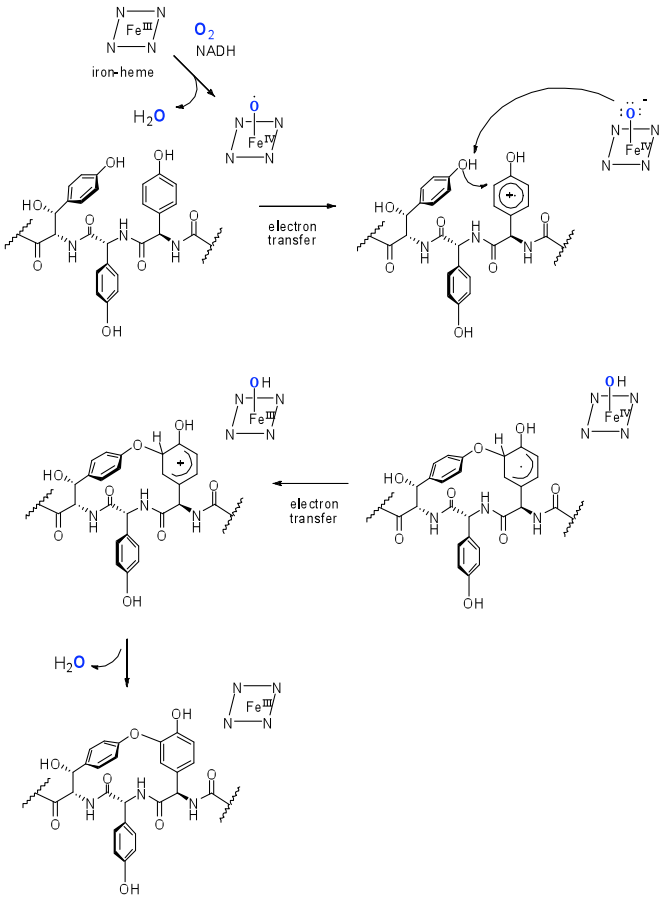


Figure 7: Mechanism of aryl-ether cross-linking enzyme OxyB.

linking enzymes. Without any optimization experiments the NRPS assembly line is able to replace an amino acid with a similar building block. OxyC is then able to use this modified side chain in the oxidative aryl-aryl coupling. These results suggest a system with high enough plasticity to undergo manipulation.

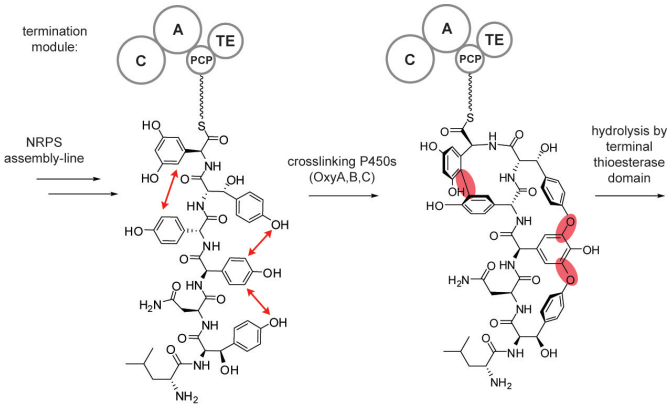


Figure 8: Chemical representation of cross-linking enzymes OxyA/B/C.

The Robinson group has shown a diverse substrate scope for OxyB.[22,23] The enzyme catalyzing the first cyclization step shows flexibility as it cyclizes a substrate with altered stereochemistry at the sixth position. Truncated vancomycin peptides are also cyclized by OxyB, specifically penta- and tri- peptide substrates. Pentapeptides with varying functional groups at the 1 position are also cyclized by OxyB.

A better understanding of the cross-linking enzymology will allow manipulation of the chemical steps that create the rigid cup shape of vancomycin that is pivotal in its biological activity.

Conclusion

Beginning with the isolation and characterization of the vancomycin family of antibiotics, a great deal has been learned about these structurally complex natural products. From sequencing the gene clusters to structurally characterizing the enzymes

from vancomycin family pathways, a large body of work has been achieved in this field. In the area of oxidation chemistry interesting cases of selectivity in addition to enzymes that perform chemistry without the use of cofactors or metals have been illuminated. The study of vancomycin enzymology has expanded our knowledge of chemistry possible in nature. The continuing study of systems that produce complex natural products will further surprise us as novel reactivity is discovered.

References:

1 Hubbard, B. K. and Walsh, C. T., Vancomycin assembly: nature's way. *Angew Chem Int Ed Engl* **42** (7), 730 (2003).

2 Cane, D. E., Walsh, C. T., and Khosla, C., Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282** (5386), 63 (1998).

3 Stephen J. Hammond, Michael P. Williamson, Dudley H. Williams, LaVerne D. Boeck, and Gary G. Marconi, On the Biosynthesis of the Antibiotic Vancomycin. *J. Chem. Soc., Chem. Commun.*, 344 (1982).

4 Lindblad, B., Lindstedt, G., Lindstedt, S., and Rundgren, M., Purification and some properties of human 4-hydroxyphenylpyruvate dioxygenase (I). *J Biol Chem* **252** (14), 5073 (1977).

5 Crouch, Nicholas P. et al., A mechanistic rationalisation for the substrate specificity of recombinant mammalian 4-hydroxylphenylpyruvate dioxygenase (4-HPPD). *Tetrahedron* **53** (20), 6993 (1997).

6 Choroba, Oliver W., Williams, Dudley H., and Spencer, Jonathan B., Biosynthesis of the Vancomycin Group of Antibiotics: Involvement of an Unusual Dioxygenase in the Pathway to (S)-4-Hydroxyphenylglycine. *Journal of the American Chemical Society* **122**, 5389 (2000).

7 Hubbard, B. K., Thomas, M. G., and Walsh, C. T., Biosynthesis of L-p-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. *Chem Biol* **7** (12), 931 (2000).

8 Neidig, M. L. et al., Spectroscopic and electronic structure studies of aromatic electrophilic attack and hydrogen-atom abstraction by non-heme iron enzymes. *Proc Natl Acad Sci U S A* **103** (35), 12966 (2006).

9 Brownlee, J. M., Johnson-Winters, K., Harrison, D. H., and Moran, G. R., Structure of the ferrous form of (4-hydroxyphenyl)pyruvate dioxygenase from Streptomyces avermitilis in complex with the therapeutic herbicide, NTBC. *Biochemistry* **43** (21), 6370 (2004).

10 Serre, L. et al., Crystal structure of Pseudomonas fluorescens 4-hydroxyphenylpyruvate dioxygenase: an enzyme involved in the tyrosine degradation pathway. *Structure* **7** (8), 977 (1999).

11 Gunsior, M., Ravel, J., Challis, G. L., and Townsend, C. A., Engineering p-hydroxyphenylpyruvate dioxygenase to a p-hydroxymandelate synthase and evidence for the proposed benzene oxide intermediate in homogentisate formation. *Biochemistry* **43** (3), 663 (2004).

12 Brownlee, J., He, P., Moran, G. R., and Harrison, D. H., Two roads diverged: the structure of hydroxymandelate synthase from Amycolatopsis orientalis in complex with 4-hydroxymandelate. *Biochemistry* **47** (7), 2002 (2008).

13 Denisov, I. G., Makris, T. M., Sligar, S. G., and Schlichting, I., Structure and chemistry of cytochrome P450. *Chem Rev* **105** (6), 2253 (2005).

14 Chen, H. and Walsh, C. T., Coumarin formation in novobiocin biosynthesis: beta-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI. *Chem Biol* **8** (4), 301 (2001).

15 Bischoff, D. et al., The Biosynthesis of Vancomycin-Type Glycopeptide Antibiotics-New Insights into the Cyclization Steps This work was supported by the Deutsche Forschungsgemeinschaft (SFB 323). We thank M. Schierle, Dr. S. Stevanovic and Prof. H.-G. Rammensee for help with Edman degradation and J. Turner, Prof. B. List and Prof. D. Boger (La Jolla, USA) for discussions on the work. *Angew Chem Int Ed Engl* **40** (9), 1693 (2001).

16 Bischoff, D. et al., The Biosynthesis of Vancomycin-Type Glycopeptide Antibiotics-The Order of the Cyclization Steps This work was supported by the Deutsche Forschungsgemeinschaft (SFB 323) and by a grant of the EU (MEGATOP, QLK3-1999-00650). R. D. S. gratefully acknowledges the support of a Feodor-Lynen Fellowship granted by the Alexander-von-Humboldt Stiftung. We thank Corina Bihlmaier and Volker Pfeifer for help with transformation and Southern hybridization, J. A. Moss (La Jolla (USA)) for critical comments on the manuscript and Prof. Dr. M. E. Maier and Prof. Dr. H.-P. Fiedler (Tubingen) for generous support. *Angew Chem Int Ed Engl* **40** (24), 4688 (2001).

17 Zerbe, K. et al., Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. *J Biol Chem* **277** (49), 47476 (2002).

18 Pylypenko, O. et al., Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C-C coupling reaction during vancomycin biosynthesis. *J Biol Chem* **278** (47), 46727 (2003).

19 Zerbe, K. et al., An oxidative phenol coupling reaction catalyzed by oxyB, a cytochrome P450 from the vancomycin-producing microorganism. *Angew Chem Int Ed Engl* **43** (48), 6709 (2004).

20 Woithe, K. et al., Oxidative phenol coupling reactions catalyzed by OxyB: a cytochrome P450 from the vancomycin producing organism. implications for vancomycin biosynthesis. *J Am Chem Soc* **129** (21), 6887 (2007).

21 Bischoff, D. et al., The biosynthesis of vancomycin-type glycopeptide antibiotics--a model for oxidative side-chain cross-linking by oxygenases coupled to the action of peptide synthetases. *Chembiochem* **6** (2), 267 (2005).

22 Geib, N. et al., New insights into the first oxidative phenol coupling reaction during vancomycin biosynthesis. *Bioorg Med Chem Lett* **18** (10), 3081 (2008).

23 Woithe, K. et al., Exploring the substrate specificity of OxyB, a phenol coupling P450 enzyme involved in vancomycin biosynthesis. *Org Biomol Chem* **6** (16), 2861 (2008).